# MONTCAS, PHASE 2

# Criterion-Referenced Test

# 2005

# TECHNICAL MANUAL

OPI

Linda McCulloch, Superintendent

Montana Office of Public Instruction
PO Box 202501
Helena, Montana 59620-2501
www.opi.state.mt.us

# TABLE OF CONTENTS

# SECTION I: ASSESSMENT DEVELOPMENT

## CHAPTER 1—BACKGROUND AND OVERVIEW

### PURPOSE OF THIS MANUAL

In the spring of 2005, Montana students in grades 4, 8, and 10 participated in the MontCAS, Phase 2 Criterion Referenced Test (CRT) in reading and mathematics in order to measure their reading and mathematics achievement as articulated by the Montana content and performance standards. This represents the second year of the CRT program, which will expand during the next two years to include additional grades and subject areas.

The purpose of this manual is to describe several technical aspects of the CRT in an effort to contribute to the accumulation of validity evidence to support CRT score interpretations. Because it is the interpretations of test scores that are evaluated for validity, not the test itself, this manual presents documentation to substantiate intended interpretations (AERA, 1999). Subsequent chapters of this manual discuss test development, test alignment, test administration, scoring, equating, item analyses, reliability, scaled scores, performance levels and reporting. Each of these topics contributes important information to the validity argument. However, note that certain aspects of a comprehensive validity argument are not included in this report, but could also be important to consider when drawing conclusions about validity. Additional sources of validity evidence might speak to the extent to which scores from the CRT assessments converge with other measures of the same or similar constructs and diverge from measures of different constructs, as well as additional consequences arising from scores at the student, school, district and state levels.

Historically, while some parts of a technical report may have been used by educated laypersons, the intended audience was experts in psychometrics and educational research. This edition of the CRT technical report is a first attempt to make the information contained herein more accessible to educated lay people by providing richer descriptions of general categories of information. In making some of the information more accessible we have purposefully preserved the depth of technical information that has historically been provided in our technical manuals. The reader will find that some of the discussion and tables continue to require a working knowledge of measurement concepts such as "reliability" and "validity", and statistical concepts such as "correlation" and "central tendency." To fully understand some data, the reader will also have to possess basic familiarity with advanced topics in measurement and statistics.

## OVERVIEW OF THE ASSESSMENT SYSTEM

On April 5, 2002, the Montana Office of Public Instruction (OPI) entered into a compliance agreement with the U.S. Department of Education that required Montana to implement a number of actions by April 5, 2005, to bring the state into compliance with the provisions of the following federal laws: Title 1 of the Elementary and Secondary Education Act (ESEA) of 1994, P.L. 103-382 and the No Child Left Behind Act (NCLB) of 2001. Montana received federal appropriations to develop an appropriate assessment. The CRT was developed in accordance with the compliance agreement and federal laws.

The CRTs are based on, and aligned to, Montana's Content Standards in Reading and Mathematics. Montana educators worked with OPI and its contractor, Measured Progress, in the development and review (content and bias) of these tests to assess how well students have learned the Montana content standards for their grade. The United States Department of Education (USDOE) approved the CRT assessments in reading and mathematics for grades 3-8 and 10 by school year 2005-2006 and in science at one grade in each of three grade spans (e.g., four, eight, and ten) by school year 2007-2008.

CRT scores are intended to be useful indicators of the extent to which students have mastered material outlined in the Montana reading and mathematics content standards. For a particular student, his/her CRT score should be used as part of a body of evidence regarding mastery and should not be used in isolation to make high stakes decisions. CRT scores, when aggregated to school, system or state levels, are more reliable indicators of program success, particularly when monitored over the course of several years.

## OPTIONS FOR PARTICIPATION

All Montana students enrolled in accredited schools are expected to participate in either the CRT or the CRT Alternate assessment. The vast majority of students will participate in the CRT, and most of them will participate under standard administration procedures. However, there is an array of standard accommodations which are available to any student, with or without disabilities, when such accommodations are necessary to allow the student to demonstrate his/her skills and competencies. Standard accommodations do not change the construct being measured and may be provided to students for either the reading or math portions of the assessment, or both, as necessary. Student's tests are scored the same way regardless of whether or not they took the test using standard accommodations.

In addition to standard accommodations, other accommodations for the CRT are available to a student when specified in his/her IEP, 504, or LEP plan. These other accommodations are referred to as non-standard accommodations and, because they alter the construct being measured, affect the student's score on the CRT. When a non-standard accommodation is used, the student's score will be reported as the lowest possible score (i.e., a scaled score of 200 which falls into the Novice performance level) for that content area. Non-standard accommodations on the CRT may be provided in reading or math, or both, as dictated by the student's IEP, 504, or LEP plan.

For a very small percentage of students, participation in the statewide assessment program will be achieved by participating in the CRT Alternate assessment. Students with significant cognitive disabilities who are working toward alternate academic achievement standards, as documented in their IEP plans, are eligible to take the CRT Alternate assessment. Technical characteristics of the CRT Alternate assessment program are described in a companion technical manual.

## BRIEF SUMMARY OF TECHNICAL EVIDENCE IN THIS MANUAL

The *Standards for Educational and Psychological Testing* (1999) provides a framework for describing sources of evidence that should be considered when constructing a validity argument. These sources include evidence based on the following five general areas: test content, response processes, internal structure, relationship to other variables, and consequences of testing. Although each of these sources may speak to a different *aspect* of validity, they are not distinct *types* of validity. Instead, each contributes to a body of evidence about the comprehensive validity of score interpretations.

Viewed through this lens provided by the Standards, evidence based on test content is extensively described in Chapters 2 through 6. Item alignment with Montana content standards; item bias, sensitivity and content appropriateness review processes; adherence to the test blueprint; use of multiple item types; use of standardized administration procedures, with accommodated options for participation; and appropriate test administration training are all components of validity evidence based on test content.

The scoring information in Chapter 7 presents evidence based on response processes and describes the steps taken to train and monitor hand-scorers, as well as quality control procedures related to scanning and machine scoring.

Evidence based on internal structure is presented in great detail in the discussions of item analyses in Chapter 8. Technical characteristics of the internal structure of the assessments are presented in terms of classical item statistics (item difficulty, item-test correlations), differential item functioning analyses, a variety of reliability coefficients, standard errors of measurement, and item response theory parameters and procedures.

Evidence based on the consequences of testing is addressed in the scaled scores, equating, and reporting information in Chapters 10 and 11, as well as in the test interpretation guide, which is a separate document that is referenced in the discussion of reporting. Each of these chapters speaks to the efforts undertaken to promote accurate and clear information provided to the public regarding test scores.

With this introduction to a conceptual understanding of how the information presented in this manual contributes to an overarching validity argument in mind, the reader should be in position to organize the extensive detail contained in the following chapters. The organization of this manual is based on the conceptual flow of an assessment cycle. The manual begins with the initial test specification and addresses all the intermediate steps that lead to final score reporting.

# CHAPTER 2—OVERVIEW OF TEST DESIGN

## CRT

Items on the CRT originate from the Progress Toward Standards (PTS) and Montana-augmented item banks (see Chapter 3) and are directly linked to **Montana's Content Standards**. The content standards are the basis for the reporting categories developed for each subject area and are used to help guide the development of test items. No other content or process is subject to statewide assessment. An item may address part, all, or several of the benchmarks within a standard.

## ITEM TYPES

Montana's educators and students were familiar with most of the item types that were used in the assessment program. The types of items used and the functions of each are described below.

**Multiple-choice items** were used, in part, to provide breadth of coverage of a content area. Because they require no more than a minute for most students to answer, these items make efficient use of limited testing time and allow coverage of a wide range of knowledge and skills.

**Short-answer items** were used to assess students' skills and their abilities to work with brief, well-structured problems that had one or a very limited number of solutions (e.g., mathematical computations). Short-answer items require approximately two minutes for most students to answer. The advantage of this type of item is that it requires students to demonstrate knowledge and skills by generating, rather than merely selecting, an answer.

**Constructed-response items** typically require students to use higher-order thinking skills—evaluation, analysis, summarization, and so on—in constructing a satisfactory response. Constructed-response items should take most students approximately five to ten minutes to complete. It should be noted that the use of released CRT items to prepare students to answer this kind of item is appropriate and encouraged.

## COMMON-MATRIX DESIGN

The CRT measures what students know and are able to do by using a variety of item types. The tests are structured using both **common** and **matrix-sampled** items. Common items are those taken by all students at a given grade level. Students' scores are based only on common items. In addition, a larger pool of matrix-sampled items is divided among the sixteen forms of the test at each grade level. Each student takes only one form of the test and so answers a fraction of the matrix-sampled items in the entire pool. The matrix-sampled items (field test items) were transparent to test takers and had a negligible impact on testing time. Because the field test was taken by all students, it provided the sample size needed to produce reliable data (750-1500 students per item as some items were repeated across forms) on which to inform item selection for future tests.

The CRT reports were delivered to schools on June 3, 2005. In addition, common items were released with a data management tool called *iAnalyze* (see Chapter 12: "Reporting".)

# CHAPTER 3—TEST DEVELOPMENT PROCESS

## CRT ITEM DEVELOPMENT

As previously mentioned, items in the CRT are derived from either the Progress Toward Standards (PTS) item bank or Montana-augmented item bank. The item development process for both item banks is similar and is discussed in greater detail in this chapter.

## PTS ITEM DEVELOPMENT

The items developed for the Progress Toward Standards (PTS) common and matrix item bank and forms were consistent with the PTS Content Standards. Measured Progress development specialists then aligned the items to the appropriate Montana Content Standards. As an additional quality control check, lead developers in each content area and Montana educators verified that each item was appropriately aligned. In January 2005, Northwest Regional Educational Laboratory (NWREL) performed an independent alignment study to verify item alignment to Montana Content Standards of both the PTS items and the Montana-specific (augmented) items.

The development process Measured Progress followed combined the expertise of the item development team and a panel of educators nationwide to help ensure that these items met the needs of the core PTS program and the CRT program. All items used in the PTS common and matrix portions of the CRT program underwent review by a national panel of content and bias reviewers. This panel included numerous Montana educators (see Appendix A: PTS Item and Bias Review Committees and Guidelines for PTS Reading Passage & Item Bias and Sensitivity Review). Annual PTS item development is depicted in the following tables:

TABLE 3-1: TOTAL NUMBER OF PTS ITEMS DEVELOPED PER YEAR

| GRADE | READING | MATH |
|---|---|---|
| 4 | 160 | 78 |
| 8 | 160 | 78 |
| 10 | 160 | 78 |

### TABLE 3-2: ANNUAL PTS READING ITEM DEVELOPMENT GRADES 4, 8 & 10

| Passages | Multiple Choice | Constructed Response |
|---|---|---|
| 2 long literary passages | 40 | 4 |
| 2 long informational passages | 40 | 4 |
| 4 short literary passages | 40 | 0 |
| 4 short informational passages | 40 | 0 |
| **12** | **160** | **8** |

### TABLE 3-3: ANNUAL PTS MATH ITEM DEVELOPMENT GRADES 4, 8 & 10

| Multiple Choice | Short Answer | Constructed Response |
|---|---|---|
| 68 | 4 | 6 |

## ITEM DEVELOPMENT PROCESS OVERVIEW

An overview of the test development process for the common and matrix items, including conducting the field tests, follows.

### TABLE 3-4: DEVELOPMENT PROCESS OVERVIEW

| DEVELOPMENT STEP | STEP DETAILS |
|---|---|
| Select reading passages and conduct external review for bias and sensitivity issues | • Measured Progress Curriculum and Assessment Specialists located potential reading passages.<br>• Reading passages were reviewed for bias and sensitivity issues before the development of reading item sets. |
| Develop items (2002 through 2004, on-going annually) | • Measured Progress Curriculum and Assessment Specialists developed reading item sets and mathematics items. |
| Review items for bias and sensitivity issues and content appropriateness (September 2002, December 2002; March 2003; December 2003; May 2004; September 2004; On-going annually) | • An external panel of educators reviewed newly-developed reading and mathematics items for bias and sensitivity issues and content appropriateness. |

| | |
|---|---|
| Edit items (2002-2004, On-going annually) | All items were reviewed by members of Measured Progress's Publications staff to assure<br>• clarity and unambiguousness of items.<br>• correct grammar, punctuation, usage, and spelling.<br>• technical quality with respect to stems, options, and scoring guides. |
| Field test items (Spring 2002 and Fall 2003) | • Measured Progress administered a field test to a sample of students in seven states prior to the first use of the items in the operational assessment. |
| Item Selection (August 2003) | • Measured Progress test developers reviewed the results of the Spring 2002 and Fall 2003 field tests and selected PTS common items for the Spring 2004 operational MontCAS forms. |
| Field test items (Spring 2004) | • Measured Progress administered a field test of the newly developed items for use in Spring 2005 as embedded matrix items on the Spring 2004 operational MontCAS forms. |
| Item Selection (August 2004) | • Measured Progress test developers reviewed the results of the Spring 2004 field test and selected PTS common items for the Spring 2005 operational MontCAS forms. |

## MONTANA-AUGMENTED ITEM DEVELOPMENT

The items developed for the augmented CRT item bank were consistent with Montana's content standards. Using a collaborative model, our development specialists worked with OPI and Montana educators to align the items developed to augment the CRT to appropriate Montana content standards. As an additional quality control check, lead developers in each content area checked for their agreement that each item was appropriately aligned. Where there were any apparent discrepancies, our lead Curriculum and Assessment specialists resolved them with OPI.

The development process Measured Progress followed, combining the expertise of the item development team and Montana educators, helped ensure that these items met the needs of the CRT program. The item specifications were built on the Montana content standards, thus assuring complete alignment between the content standards and the augmented portion of the CRT. In addition to internal review, all test materials and items used in the CRT program underwent review by Montana educators and bias review committees prior to print. Table 3-5 depicts the number of items developed and field tested in 2002-2003 to support the program's item bank 2004 through 2007.

### TABLE 3-5: TOTAL NUMBER OF MONTANA-AUGMENTED ITEMS DEVELOPED AND FIELD TESTED BY GRADE AND CONTENT (ALL MULTIPLE CHOICE ITEMS)

| GRADE | READING | MATH |
|---|---|---|
| 3 | 60 | 60 |
| 4 | 100 | 100 |
| 5 | 60 | 60 |
| 6 | 60 | 60 |
| 7 | 60 | 60 |
| 8 | 100 | 100 |
| 10 | 150 | 150 |

## MONTANA-AUGMENTED ITEM DEVELOPMENT PROCESS OVERVIEW

An overview of the test development process for the Montana-augmented item bank, including conducting the field tests (April 2003), follows.

### TABLE 3-6: DEVELOPMENT PROCESS OVERVIEW

| DEVELOPMENT STEP | STEP DETAILS |
|---|---|
| Review by Montana educators of passages for the reading tests (Aug. 2002) | • Measured Progress Curriculum and Assessment reading specialists located potential reading passages.<br>• MT educators approved the passages in consultation with a Montana Bias Review Committee prior to item writing.<br>• Measured Progress Permissions staff secured permissions to use the passages prior to item writing meetings. |
| Item drafting/editing meetings (Sept. 2002) | Measured Progress Curriculum and Assessment specialists<br>• provided item development training to Montana participants;<br>• facilitated the development of item ideas by the participants. |
| Editorial review of items (Oct. 2002) | All items were reviewed by members of Measured Progress's Publications staff to ensure<br>• clarity and unambiguousness of items;<br>• correct grammar, punctuation, usage, and spelling;<br>• technical quality with respect to stems, options, and scoring guides;<br>• compliance with OPI sensitivity standards and style guidelines. |
| Item review meetings (Nov. 2002) | Curriculum and Assessment Specialists facilitated the review of all items with Montana educators and selected appropriate items for field testing in 2003. |
| Bias Review Committee meetings (Nov. 2002) | Measured Progress staff facilitated the review of all test items for sensitivity and bias considerations based on OPI guidelines. Members of this committee were selected by OPI. Measured Progress provided OPI with guidelines for committee membership. |
| Field Test of MT-Augmented Items (April 2003) | Measured Progress provided field test forms which were administered to a sample of students in Montana prior to use of the items in operational assessment to assure quality of items. |
| Final Item Selection (August 2003) | Measured Progress provided the reports necessary for Montana educators to review the results of field-testing, revise as necessary, and select items for the augmented portion of the assessment. |

## INTERNAL ITEM REVIEW

The lead or peer Curriculum and Assessment Specialist within the content specialty reviewed each item for:

- item “integrity”, item content and structure, appropriateness to designated content area, item format, clarity, possible ambiguity, keyability, single “keyness”, appropriateness and quality of reading selections and graphics, and appropriateness of scoring guide descriptions and distinctions (as correlated to the item and within the guide itself).
- scorability and evaluated whether the scoring guide adequately addressed performance on the item.
- fundamental issues including the following:
  - What is the item asking?
  - Is the key the only possible key?
  - Is the constructed-response item scorable as written (are the correct words used to elicit the response defined by the guide)?
  - Is the wording of the scoring guide appropriate and parallel to the item wording?
  - Is the item complete (i.e., with scoring guide, content codes, key, grade level, and contract identified)?
  - Is the item appropriate for the designated grade level?

## EXTERNAL ITEM AND BIAS REVIEWS

All PTS and Montana-augmented items undergo the following external reviews:

- In October 2004, common item sets were delivered to OPI for Montana educator content and bias reviews. Feedback from the content and bias reviews was incorporated into the final editing processes.
- The PTS National Bias and Content Review Committee reviewed the common and matrix passages and items used for the 2005 administration in Montana during two two-day meetings, held in March 2003 and December 2003 in Chicago, IL, and during a mail review of passages in July 2003 (see Appendix A).

## ITEM EDITING

Editors reviewed and edited the items to ensure uniform style (based on *The Chicago Manual of Style,* 14th edition) and adherence to sound testing principles. These principles included the stipulation that items

- were correct with regard to grammar, punctuation, usage, and spelling;
- were written in a clear, concise style;
- contained unambiguous explanations for students as to what was required to attain a maximum score;
- were written at a reading level that would allow the student to demonstrate his or her knowledge of the tested subject matter regardless of reading ability;
- exhibited high technical quality regarding psychometric characteristics;
- had appropriate answer options or score-point descriptors; and
- were free of potentially insensitive content.

## OPERATIONAL TEST ASSEMBLY

Test assembly is the sorting and laying out of item sets into test forms. Criteria considered during this process included the following:

- **Content coverage/match to test design**. The curriculum specialist completed an initial sorting of items into sets based on a balance of content categories across sessions and forms, as well as a match to the test design (e.g., number of multiple-choice, short-answer, and constructed-response items).
- **Item difficulty and complexity**. Item statistics drawn from the data analysis of previously tested items were used to ensure that there were similar levels of difficulty and complexity across forms.
- **Visual balance**. Item sets were reviewed to ensure that each reflected a similar length and "density" of selected items (e.g., length/complexity of reading selections or number of graphics).
- **Option balance**. Each item set was checked to verify that it contained a roughly equivalent number of key options (As, Bs, Cs, and Ds).
- **Name balance**. Item sets were reviewed to ensure that a diversity of names was used.
- **Bias**. Each item set was reviewed to ensure fairness and balance based on gender, ethnicity, religion, socioeconomic status, and other factors.

- **Page fit**. Item placement was modified to ensure the best fit and arrangement of items on any given page.
- **Facing-page issues**. For multiple items associated with a single stimulus (a graphic or a reading selection), consideration was given to whether those items needed to begin on a left- or right-hand page, as well as to the nature and the amount of material that needed to be placed on facing pages. These considerations served to minimize the amount of page flipping required of the students.
- **Relationships between forms**. Sets of common items were placed identically in each version of the forms. Although matrix-sampled item sets differed from form to form, they took up the same number of pages in each form so that sessions and content areas began on the same page in every form. Therefore, the number of pages needed for the longest form often determined the layout of each form.
- **Visual appeal**. The visual accessibility of each page of the form was always taken into consideration, including such aspects as the amount of white space, the density of the text, and the number of graphics.

## EDITING DRAFTS OF OPERATIONAL TESTS

Any changes made during the test construction had to be reviewed and approved by the Curriculum and Assessment Specialist. Once a form had been laid out in what was considered its final form, it was reread to identify any final considerations, including the following:

- **Editorial changes**. All text was scrutinized for editorial accuracy, including consistency of instructional language, grammar, spelling, punctuation, and layout. Measured Progress's publishing standards are based on *The Chicago Manual of Style*.
- **Keying items**. Items were reviewed for any information that might "key" or provide information that would help students answer another item. Decisions about moving keying items were based on the severity of the key-in and the placement of the items in relation to each other within the form.
- **Key patterns**. The final sequence of keys was reviewed to ensure that the order appeared random (i.e., no recognizable pattern and no more than three of the same key in a row).

## BRAILLE AND LARGE-PRINT TRANSLATION

Form one for grades 4, 8, and 10 tests was translated into Braille by National Braille Press, a subcontractor that specializes in test materials for blind and visually handicapped students. In addition, form one for each grade was adapted into a large-print version.

# CHAPTER 4—DESIGN OF THE READING ASSESSMENT

## READING TEST BLUEPRINT

As indicated earlier, the test blueprint for reading was based on PTS and Montana's reading content standards, which identifies five **Montana Content Standards** that apply specifically to reading and reading comprehension. Those content standards follow:

- **Reading Standard 1:** Students construct meaning as they comprehend, interpret, and respond to what they read.
- **Reading Standard 2:** Students apply a range of skills and strategies to read.
- **Reading Standard 3:** Students set goals, monitor, and evaluate their reading progress. (Cannot measure this benchmark with traditional paper/pencil test.)
- **Reading Standard 4:** Students select, read, and respond to print and nonprint material for a variety of purposes.
- **Reading Standard 5:** Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences.

### TABLE 4-1: GRADES 3-8 READING TEST BLUEPRINT

| READING GRADES 3-8 (PER FORM) | |
|---|---|
| **Passages** | **Number of items** |
| Session 1 Common | |
| Short passage | 5 MC |
| Short passage | 5 MC |
| Long passage | 11 MC, 1 CR |
| **Session total** | **21 MC, 1 CR** |
| Session 2 Montana-specific common and embedded matrix field test | |
| Montana-specific passage (common) | 10 MC |
| Embedded long passage (field test) | 6 MC, 1 CR |
| Embedded short passage (field test) | 6 MC |
| **Session total** | **22 MC, 1 CR** |
| Session 3 Common | |
| Short passage | 5 MC |
| Short passage | 5 MC |
| Long passage | 11 MC, 1 CR |
| **Session total** | **21 MC, 1 CR** |
| **Common total** | **52 MC, 2 CR** |

## TABLE 4-2: GRADE 10 READING BLUEPRINT

| READING GRADE 10 (PER FORM) | |
|---|---|
| **Passages** | **Number of items** |
| Session 1 Common | |
| Short passage | 5 MC |
| Short passage | 5 MC |
| Long passage | 11 MC, 1 CR |
| **Session total** | **21 MC, 1 CR** |
| Session 2 Montana-specific common and embedded matrix (field test) | |
| Montana-specific passage (common) | 15 MC |
| Embedded long passage (field test) | 6 MC, 1 CR |
| Embedded short passage (field test) | 6 MC |
| **Session total** | **27 MC, 1 CR** |
| Session 3 Common | |
| Short passage | 5 MC |
| Short passage | 5 MC |
| Long passage | 11 MC, 1 CR |
| **Session total** | **21 MC, 1 CR** |
| **Common total** | **57 MC, 2 CR** |

**Key**

- MC = multiple-choice items
- CR = constructed-response items

Passages included both long and short texts selected from reading sources that students at each grade level would be likely to encounter in their classroom and in their independent reading. No passages were written specifically for the assessment, but instead were collected from published works.

- **Literary passages** are represented by a variety of genres—modern narratives; diary entries; drama; poetry; biographies; essays; excerpts from novels; short stories; and traditional narratives, such as fables, myths, and folktales.
- **Content passages** are primarily informational and often deal with the areas of science and social studies. They are drawn from such sources as newspapers, magazines, and books.
- **Practical passages** are functional materials that instruct or advise the reader—for example, directions, reference tools, or manuals.

The main difference in the passages used for grades 4, 8, and 10 was their degree of difficulty. All passages were selected to be appropriate for the intended audience; however, the ideas expressed became increasingly more complex at grade levels 8 and 10.

The items related to these passages required students to demonstrate their skills in both literal comprehension (where the answer is stated explicitly in the text) and inferential comprehension (where the answer is implied by the text and/or the text must be connected to relevant prior knowledge to determine an answer). In addition, some items focused on the reading skills reflected in content standards. Items of this type required students to use the skills and strategies of reading to answer items—for example, how to identify the author's principal purpose, such as to persuade, entertain, or inform—and to demonstrate their understanding of how words and images communicate to readers. The table below depicts passage distribution, length, and reporting categories.

### TABLE 4-3: DISTRIBUTION

| **Reading Passage Distribution** | | | |
|---|---|---|---|
| Literary | | 50% | 25 points |
| Informational | Comprised of both content and practical passages | 50% | 25 points |
| | | 100% | 50 points |
| **Reading Passage Length** | | | |
| Long* | Either a literary or informational per session | 50% | 25 points |
| Short* | At least one literary and informational per session | 50 % | 25 points |
| | | 100% | 50 points |
| **Reporting Categories** | | | |
| Comprehension and Analysis | | 70% | 35 points |
| Reading Process and Skills | | 30 % | 15 points |
| | | 100 % | 50 points |

## ITEM TYPES

The CRT assessment in reading included multiple-choice and constructed-response items (see Table 4-4 below). Constructed-response items required students to write an answer consisting of several phrases or short sentences. Each type of item was worth a specific number of points in the student's total language arts score.

### TABLE 4-4: ITEM TYPES

| Type of Item | Possible Score Points |
|---|---|
| Multiple-Choice | 0 or 1 |
| Constructed-Response | 1, 2, 3, or 4 |

## TEST DESIGN

The table below summarizes the number and types of items that were used in the CRT reading assessment for 2005 and shows the construction of the common portions of the assessment.

### TABLE 4-5: TEST DESIGN

| | | Common Reading Items | | TOTAL | |
|---|---|---|---|---|---|
| **Grade** | **Session 1** | **Session 2** | **Session 3** | **MC** | **CRs** |
| 4 | 21 MC, 1 CR | 10 MC | 21 MC, 1 CR | 52 | 2 |
| 8 | 21 MC, 1 CR | 10 MC | 21 MC, 1 CR | 52 | 2 |
| 10 | 21 MC, 1 CR | 15 MC | 21 MC, 1 CR | 57 | 2 |

**Key**

- MC = multiple-choice items
- CR = constructed-response items

# CHAPTER 5—DESIGN OF THE MATHEMATICS ASSESSMENT

## MATHEMATICS BLUEPRINT

The mathematics framework was based on Montana's Mathematics Content Standards, which identifies seven **content standards**, as shown below:

- **Mathematics Standard 1:** Problem Solving
- **Mathematics Standard 2:** Numbers and Operations
- **Mathematics Standard 3:** Algebra
- **Mathematics Standard 4:** Geometry
- **Mathematics Standard 5:** Measurement
- **Mathematics Standard 6:** Data Analysis, Statistics, and Probability
- **Mathematics Standard 7:** Patterns, Relations, and Functions

### TABLE 5-1: MATHEMATICS BLUEPRINT

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Test Design:** | 45 multiple-choice items<br>3 1-point short-answer items<br>2 4-point constructed-response items<br>Total points: 56 | | | | | | |
| **Percent Point distribution by content strand*** | | | | | | | |
| **PTS Standards** | **Grade 3** | **Grade 4** | **Grade 5** | **Grade 6** | **Grade 7** | **Grade 8** | **Grade 10** |
| **Number and Operations** | 32% | 32% | 32% | 32% | 30% | 20% | 20% |
| **Algebra** | 20% | 20% | 20% | 20% | 20% | 29% | 27% |
| **Geometry and Measurement** | | | | | | | |
| **Geometry** | 16% | 16% | 16% | 16% | 16% | 18% | 23% |
| **Measurement** | 13% | 13% | 13% | 13% | 14% | 14% | 11% |
| **Data Analysis/Probability** | 20% | 20% | 20% | 20% | 20% | 20% | 20% |
| *Because percents are rounded to the nearest whole number, not all sums add to 100%.<br>**Note:** Geometry and Measurement comprise a single reporting category. | | | | | | | |
| | | | | | | | |
| **Point distribution by content strand** | | | | | | | |
| | **Grade 3** | **Grade 4** | **Grade 5** | **Grade 6** | **Grade 7** | **Grade 8** | **Grade 10** |
| **Number and Operations** | 18 | 18 | 18 | 18 | 17 | 11 | 11 |
| **Algebra** | 11 | 11 | 11 | 11 | 11 | 16 | 15 |
| **Geometry and Measurement** | | | | | | | |
| **Geometry** | 9 | 9 | 9 | 9 | 9 | 10 | 13 |
| **Measurement** | 7 | 7 | 7 | 7 | 8 | 8 | 6 |
| **Data Analysis/Probability** | 11 | 11 | 11 | 11 | 11 | 11 | 11 |
| **Four-point items:** Each test contains two 4-point constructed-response items. In any given year, the two items will measure two different strands. From year to year, those strands may change. | | | | | | | |

**One-point items:** There are two types of one-point items: multiple-choice and short answer items. Each test contains 45 multiple-choice items and three short-answer items. The number of one-point items per strand will vary from year to year depending on which two strands are measured by the four-point items. (The number of total points per strand is kept constant from year to year.)

## Number of 1-point items per content strand

| | **Grade 3** | **Grade 4** | **Grade 5** | **Grade 6** | **Grade 7** | **Grade 8** | **Grade 10** |
|---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| **Number and Operations** | 14 or 18 | 14 or 18 | 14 or 18 | 14 or 18 | 13 or 17 | 7 or 11 | 7 or 11 |
| **Algebra** | 7 or 11 | 7 or 11 | 7 or 11 | 7 or 11 | 7 or 11 | 12 or 16 | 11 or 15 |
| **Geometry and Measurement** | | | | | | | |
| **Geometry** | 5 or 9 | 5 or 9 | 5 or 9 | 5 or 9 | 5 or 9 | 6 or 10 | 9 or 13 |
| **Measurement** | 3 or 7 | 3 or 7 | 3 or 7 | 3 or 7 | 4 or 8 | 4 or 8 | 2 or 6 |
| **Data Analysis/Probability** | 7 or 11 | 7 or 11 | 7 or 11 | 7 or 11 | 7 or 11 | 7 or 11 | 7 or 11 |

## Distribution of One-Point Items Within Strand by Standard

The distribution of one-point items within a strand is partially dependent on the specific items selected for a given test. However, a minimal number of one-point items per standard has been established. Those numbers are shown in the table below.

## Minimum Number of 1-Point Items Per Strand

| | **Grade 3** | **Grade 4** | **Grade 5** | **Grade 6** | **Grade 7** | **Grade 8** | **Grade 10** |
|---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| **Number and Operations** | | | | | | | |
| ***Total Number of points*** | ***18*** | ***18*** | ***18*** | ***18*** | ***17*** | ***11*** | ***11*** |
| Number concepts | 4 | 3 | 2 | 3 | 3 | 2 | 2 |
| Meanings of operations | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Computation/estimation | 4 | 5 | 6 | 5 | 4 | 2 | 2 |
| *Floating points* | *5 or 9* | *5 or 9* | *5 or 9* | *5 or 9* | *5 or 9* | *2 or 6* | *2 or 6* |
| | | | | | | | |
| **Algebra** | | | | | | | |
| ***Total Number of points*** | ***11*** | ***11*** | ***11*** | ***11*** | ***11*** | ***16*** | ***15*** |
| Patterns | 3 | 2 | 2 | 1 | 1 | 1 | 1 |
| Algebraic symbols | 1 | 1 | 1 | 2 | 2 | 4 | 4 |
| Mathematical models | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Change | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Floating points* | *1 or 5* | *2 or 6* | *2 or 6* | *2 or 6* | *2 or 6* | *5 or 9* | *4 or 8* |
| | | | | | | | |
| **Geometry and Measurement** | | | | | | | |
| **Geometry** | | | | | | | |
| ***Total Number of points*** | ***9*** | ***9*** | ***9*** | ***9*** | ***9*** | ***10*** | ***13*** |
| Properties of 2-and 3-d shapes | 2 | 2 | 2 | 2 | 2 | 2 | 3 |
| Coordinate Geometry | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Transformations/symmetry | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Visualization/spatial reasoning | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Floating points* | *0 or 4* | *0 or 4* | *0 or 4* | *0 or 4* | *0 or 4* | *1 or 5* | *3 or 7* |
| | | | | | | | |
| **Measurement** | | | | | | | |
| ***Total Number of points*** | ***7*** | ***7*** | ***7*** | ***7*** | ***8*** | ***8*** | ***6*** |
| Concepts of measurement | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Techniques, tools, formulas | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Floating points* | *1 or 5* | *1 or 5* | *1 or 5* | *1 or 5* | *2 or 6* | *2 or 6* | *0 or 4* |
| | | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Data Analysis/Probability** | | | | | | | |
| ***Total Number of points*** | **11** | **11** | **11** | **11** | **11** | **11** | **11** |
| Collect/organize/display data | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| Statistical methods | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Inferences/predictions | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Probability | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Floating points* | *2 or 6* | *2 or 6* | *2 or 6* | *3 or 7* | *3 or 7* | *3 or 7* | *3 or 7* |

## CONTENT SPECS

For students to function effectively as mathematical problem solvers, they must be taught how to apply and communicate basic concepts and procedures, as well as how to do the procedures themselves.

**Content items** measure what students have been taught directly. Included in these are the basic concepts and procedural skills from all the content standards. For example, in the numbers and number sense standard and the computation standard, conceptual and procedural knowledge includes understanding of place value in our number system; the computational algorithms as applied to whole numbers, fractions, and decimals; and the concepts of ratio, proportion, and percent. In the data analysis and statistics standard, conceptual and procedural knowledge includes the ability to read charts and graphs as well as to understand concepts of averages (means, medians, and modes) and the methods for computing them. Contextual settings used in items measuring this category were very simple and were directly related to those used in the teaching of the concepts and the procedures.

**Application items** measure what the students can do with what they have been taught. Included are items requiring students to combine the basic concepts and procedures to solve real-life and mathematical problems, to evaluate their own ideas and the ideas of others using mathematical reasoning, and to communicate their ideas using the wealth of symbolic, pictorial, graphic, and verbal representations available in mathematics.

It is important to understand that application items also measure mastery of the basic concepts and procedures. For example, in mathematics, items were either short-answer or constructed-response items (see “Item Types” in the table below), which were worth up to four score points. In most cases, portions of these items required the student to perform some problem solving, reasoning, and/or communicating. At the same time, however, the items required the students to demonstrate their understanding of mathematics content. If a student did not show mastery of all aspects of a constructed-response item, or if he/she made careless errors, the student did not earn the highest score for that item. Thus, it can be said that **all** mathematics items in the CRT measured content; some items went beyond that realm (short-answer and constructed-response), however, and were classified as application.

### TABLE 5-2: DISTRIBUTION OF MATHEMATICS PROCESS CATEGORIES

| Grade | 3 | 4 | 5 | 6 | 7 | 8 | HS |
|---|---|---|---|---|---|---|---|
| Procedures/Concepts | 65% | 65% | 60% | 60% | 55% | 55% | 55% |
| Problem Solving/ Reasoning | 35% | 35% | 40% | 40% | 45% | 45% | 45% |

## ITEM TYPES

The CRT mathematics assessment included multiple-choice, short-answer, and constructed-response items. Short-answer items required students to perform a computation or solve a simple problem. Constructed-response items were more complex, requiring 8-10 minutes of response time. Each type of item was worth a specific number of points in the student's total mathematics score, as shown below.

### TABLE 5-3: ITEM TYPES

| Type of Item | Possible Score Points |
|---|---|
| Multiple-Choice | 0 or 1 |
| Short-Answer | 0 or 1 |
| Constructed-Response | 1, 2, 3, or 4 |

## TEST DESIGN

Table 5-4 summarizes the number and types of items that were used in the CRT mathematics assessment for 2005, and shows the construction of the common portions of the assessment.

### TABLE 5-4: TEST DESIGN

| | | Common Math Items | | | TOTAL | |
|---|---|---|---|---|---|---|
| **Grade** | **Session 1 Cal** | **Session 2A Cal** | **Session 2B No Cal** | **Session 3 No Cal** | **MC** | **SA & CRs** |
| 4 | 24 MC, 1 CR | 5 MC | 5 MC | 21 MC, 3 SA, 1 CR | 55 | 3 SA, 2 CRs |
| 8 | 24 MC, 1 CR | 5 MC | 5 MC | 21 MC, 3 SA, 1 CR | 55 | 3 SA, 2 CRs |
| 10 | 24 MC, 1 CR | 8 MC | 7 MC | 21 MC, 3 SA, 1 CR | 60 | 3 SA, 2 CRs |

**Key**

- Cal = calculator use allowed
- No Cal = no calculator use allowed
- MC = multiple-choice items
- SA = short-answer items
- CR = constructed-response items

## THE USE OF CALCULATORS IN THE CRT

The Montana educators who helped develop the CRT acknowledged the importance of mastering arithmetic algorithms. At the same time, they understood that the use of calculators is a necessary and important skill in society today. Calculators can save time and prevent error in the measurement of some higher-order thinking skills and allow students to do more sophisticated and intricate problems. For these reasons, calculators were permitted on some parts of the CRT mathematics assessment and prohibited on others. (Students were allowed to use any calculator with which they were familiar.)

# SECTION II: TEST ADMINISTRATION

## CHAPTER 6—TEST ADMINISTRATION

### RESPONSIBILITY FOR ADMINISTRATION

As indicated in the *Test Coordinator's Manual*, principals and/or their designated School Test Coordinators were responsible for the proper administration of the CRT. This manual was used to ensure the uniformity of administration procedures from school to school.

### PROCEDURES

School Test Coordinators were instructed to read the *Test Coordinator's Manual* prior to testing, and to be familiar with the instructions given in the *Test Administrator's Manual*. The *Test Coordinator's Manual* provided each school with checklists to help prepare for testing. The checklists outlined tasks to be performed before, during, and after test administration. Along with providing these checklists, the *Test Coordinator's Manual* outlined the nature of the testing material being sent to each school, how to inventory the material, how to track it during administration, and how to return the material once testing was complete. It also contained information about including or excluding students. The *Test Administrator's Manual* included checklists for the administrators to prepare themselves, their classrooms, and their students for the administration of the test. The *Test Administrator's Manual* contained sections that detailed the procedure to be followed for each test session, and it contained instructions on preparing the material prior to giving it to the School Test Coordinator for its return to Measured Progress.

### ADMINISTRATOR TRAINING

In addition to distributing the *Test Coordinator's Manuals* and *Test Administrator's Manuals*, OPI and Measured Progress conducted preadministration workshops on February 8, 2005 (one MetNet and one videostream) to train and inform school personnel about the new CRT. Training materials and the PowerPoint presentation were posted on OPI's Web site.

## PARTICIPATION REQUIREMENTS

All students were expected to participate; however, scores of students in the following categories were excluded from the calculation of averages:

- Foreign exchange students
- Students not enrolled in an accredited Montana school (for example: homeschooled student)
- Students enrolled in a private accredited school
- Students enrolled in a private nonaccredited school
- Students enrolled in a private nonaccredited Title 1 school
- Students enrolled part-time (less than 180 hours) taking a mathematics or reading course
- First year in US LEP students **were required** to participate in the math assessment only.

### TABLE 6-1: SUMMARY OF ELIGIBILITY FOR EXCLUSION FROM THE CRT

| EXCLUDED FROM AVERAGES | MUST PARTICIPATE | MAY PARTICIPATE |
|---|---|---|
| FOREIGN EXCHANGE STUDENTS | **YES** | |
| STUDENTS NOT ENROLLED IN AN ACCREDITED MONTANA SCHOOL | | **YES** |
| STUDENTS ENROLLED IN A PRIVATE ACCREDITED SCHOOL | **YES** | |
| STUDENTS ENROLLED IN A PRIVATE NONACCREDITED SCHOOL | | **YES** |
| STUDENTS ENROLLED IN A PRIVATE NONACCREDITED TITLE I SCHOOL | | **YES** |
| STUDENTS ENROLLED PART-TIME (LESS THAN 180 HRS.) TAKING A MATHEMATICS OR READING COURSE | | **YES** |
| READING: FIRST YEAR IN US LEP STUDENTS | | **YES** |
| MATHEMATICS: FIRST YEAR IN US LEP STUDENTS | **YES** | |

Information about the exclusion was coded in by staff after testing was completed. The *Test Coordinator's Manual* and *Test Administrator's Manual* provided directions on coding. Please refer to Appendix G: Reporting Decision Rules.

## TEST SCHEDULING

The CRTs were given during the spring: **reading** and **mathematics** were administered to grades 4, 8 and 10 during a four-week period (March 7–30, 2005). Schools were able to schedule testing sessions at any time during this period, provided they followed the sequence in the scheduling guidelines detailed in test administration manuals. Schools were asked to schedule makeup testing of students who were absent from initial test sessions during this testing window.

The CRT is an untimed assessment; however, guidelines or ranges were provided in the *Test Coordinator's Manual* and *Test Administrator's Manual* based on estimates of the time it would take an average student to respond to each type of item that made up the test:

- multiple-choice items – 1 minute per item
- short-answer items – 2 minutes per item
- constructed-response items – 10 minutes per item

While the guidelines for scheduling were based on the assumption that most students would complete the test within the time estimated, each test administrator was asked to allow additional time for students who needed it. If additional classroom space was not available for students who required additional time to complete the tests, schools were encouraged to consider using another space, such as the guidance office, for this purpose. If additional areas were not available, it was recommended that each classroom being used for test administration be scheduled for the maximum amount of time.

### TABLE 6-2: GRADES 4 & 8 RECOMMENDED READING SCHEDULE

| **Grades 4 & 8 Recommended Testing Schedule—Reading** | | |
|---|---|---|
| **DAY 1**<br>**Reading** | **Test Activity** | Time Range<br>(in minutes) |
| | General Instructions | 5-10 |
| | | |
| Session 1 | Reading Session 1 | 45-55 |
| **DAY 2**<br>**Reading** | | |
| Session 2 | Reading Session 2 | 45-55 |
| | Break | |
| Session 3 | Reading Session 3 | 45-55 |

### TABLE 6-3: GRADES 4 & 8
### RECOMMENDED MATHEMATICS SCHEDULE

| **Grades 4 & 8 Recommended Testing Schedule—Mathematics** | | |
|---|---|---|
| **DAY 3**<br>**Mathematics** | **Calculators ARE allowed** | Time Range<br>(in minutes) |
| Session 1 | Mathematics Session 1 | 45-55 |
| | Break | |
| Session 2A | Mathematics Session 2A | 20-30 |
| **DAY 4**<br>**Mathematics** | **Calculators are NOT allowed** | |
| Session 2B | Mathematics Session 2B | 20-30 |
| | Break | |
| Session 3 | Mathematics Session 3 | 45-55 |

### TABLE 6-4: GRADE 10
### RECOMMENDED READING SCHEDULE

| **Grade 10 Recommended Testing Schedule—Reading** | | |
|---|---|---|
| **DAY 1**<br>**Reading** | **Test Activity** | Time Range<br>(in minutes) |
| | General Instructions | 10-20 |
| | Break | |
| Session 1 | Reading Session 1 | 50-60 |
| **DAY 2**<br>**Reading** | | |
| Session 2 | Reading Session 2 | 50-60 |
| | Break | |
| Session 3 | Reading Session 3 | 50-60 |

## TABLE 6-5: GRADE 10
## RECOMMENDED MATHEMATICS SCHEDULE

| Grade 10 Recommended Testing Schedule—Mathematics | | |
|---|---|---|
| **DAY 3<br>Mathematics** | **Calculators ARE allowed** | Time Range<br>(in minutes) |
| Session 1 | Mathematics Session 1 | 50-60 |
| | Break | |
| Session 2A | Mathematics Session 2A | 20-30 |
| **DAY 4<br>Mathematics** | **Calculators are NOT allowed** | |
| Session 2B | Mathematics Session 2B | 20-30 |
| | Break | |
| Session 3 | Mathematics Session 3 | 50-60 |

# SECTION III: DEVELOPMENT AND REPORTING OF SCORES

## CHAPTER 7—SCORING

### MACHINE-SCORED ITEMS

Once the 2005 test booklets had been logged in, identified with appropriate scannable, preprinted school information sheets, examined for extraneous materials, and batched, they were moved into the scanning area. For all student response booklets (and other forms that required imaging/scanning) this was the last step in the processing loop in which the documents themselves were handled.

At that point, 100 percent of the student response documents and other scannable information necessary to produce the required reports had been captured and converted into an electronic format, including all student identification and demographics, and digital image clips of short-answer and constructed-response student responses. The digital image clip information allowed Measured Progress to replicate student responses on the readers' monitors just as they had appeared on the originals. From that point on, the entire process—data processing, scoring, benchmarking data analysis, and reporting—was accomplished without further reference to the originals.

The first step in that conversion was the removal of the booklet bindings so that the individual pages could pass through the scanners one at a time. Once cut, the sheets were put back in their proper boxes and placed in storage until needed for the scanning/imaging process.

Customized scanning programs for all scannables were prepared to selectively read the student response booklets and to format the scanned information electronically according to predetermined requirements. Any information (including multiple-choice response data) that had been designated time-critical or process-critical was handled first.

In addition to numerous real-time quality control checks, duplex read, a transport printer that prints a unique identifying number on each sheet of each booklet, and on-line editing capability, the 5000i scanners offer features that make them compatible with Internet technology.

## SCANNING QUALITY CONTROL

NCS scanners are equipped with many built-in safeguards that prevent data errors. The scanning hardware is continually monitored for conditions that will cause the machine to shut down if standards are not met. It will display an error message and prevent further scanning until the condition is corrected. The areas monitored include document page and integrity checks, user-designed on-line edits, and many internal checks of electronic functions.

Before every scanning shift begins, Measured Progress operators perform a daily diagnostic routine. This is yet another step to protect data integrity and one that has been done faithfully for the many years that we have been involved in production scanning. In the rare event that the routine detects a photocell that appears to be out of range, we calibrate that machine and perform the test again. If the read is still not up to standard, we call for assistance from our field service engineer.

As a final safeguard, spot checks of scanned files, bubble by bubble and image by image, were routinely made throughout scanning runs. The result of these precautions, from the original layout of the scanning form to the daily vigilance of our operators, was a scan error rate well below 1 per 1000.

## ELECTRONIC DATA FILES

Once the data had been entered and the scanning logs and other paperwork completed, the booklets themselves were put into storage (where they stayed for at least 180 days beyond the close of the fiscal year). When it had been determined that the files were complete and accurate, those files were duplicated electronically and made available for many other processing options. Completed files were loaded onto our local area network (LAN) for transfer to Measured Progress's proprietary I-Score system for scoring. Those files were then used to identify (and print out) papers to be used in the benchmarking processes, and the data made transferable via the Internet, CD-ROM, or optical disk.

## ITEMS SCORED BY READERS

Test and answer materials were handled as little as possible to minimize the possibility of loss, mishandling, or breach of security. Once scanned, either by optical mark reader or the I-Score system, papers were stored securely in areas with limited personnel access.

As explained in the following sections on scoring, the I-Score system itself ensures the security of responses and test items: all scoring is "blind"; that is, no student names are associated with viewed responses or raw scores and all scoring personnel are subject to the same nondisclosure requirements and supervision as regular Measured Progress staff.

## I-SCORE

After the 2005 test material had been loaded into the LAN, I-Score sent electronically scanned images of student work to individual readers at computer terminals, who evaluated each response and recorded each student's score via keypad or mouse entry. When the reader had finished with one response, the next response appeared immediately on the computer screen. In that way, the system guaranteed complete anonymity of individual students and ensured the randomization of responses during scoring.

Although I-Score is based on conventional scoring techniques, it also offers numerous benefits, not the least of which is raising the bar on scoring process capability. Some of the benefits are

- real-time information on scorer reliability, read-behinds, and overall process monitoring;
- early access to subsets of data for tasks such as standard setting;
- reduced material handling, which not only saves time and labor, but also enhances the security of materials; and
- immediate access to samples of student responses and scores for reporting and analysis through electronic media.

Scoring operations, directed by the manager of scoring services, were carried out by a highly qualified staff. The staff included

- chief readers, who oversaw all training and scoring within particular subject areas;
- quality assurance coordinators (QACs), who led benchmarking and training activities and monitored scoring consistency and rates;
- verifiers, who performed read-behinds of readers and assisted at scoring tables as necessary; and
- readers, who performed the bulk of the scoring.

The table below summarizes the qualifications of the 2005 CRT quality assurance coordinators and readers.

**TABLE 7-1: EDUCATIONAL CREDENTIALS**

| 2005 Spring Administration | | | | | |
|---|---|---|---|---|---|
| Scoring Responsibility | Educational Credentials | | | | Total |
| | Doctorate | Master's | Bachelor's | Associate's | |
| QACs | 0.00 | 53.33 | 46.67 | 0.00 | 100% |
| Readers | 4.89 | 14.66 | 39.85 | 40.60 | 100% |

## PRELIMINARY ACTIVITIES

Preliminary activities for scoring included (1) participating in the planning and design of documents to be used for scoring, (2) reviewing items and score guides for benchmarking and training and the creation of benchmarking packets, and (3) selecting scoring staff and training them for scoring.

## PLANNING AND DESIGNING DOCUMENTS

At the request of the project manager, scoring personnel advised project management and OPI staff on the program design in order to support an efficient and effective scoring process. Scoring staff also contributed to the design of

- response documents and the image-capture process to yield acceptable image clips (also defining file format and layout); and
- scoring benchmarks composed of the guide, subject background information, and anchor papers.

## BENCHMARKING

Before the scheduled start of scoring activities, scoring center staff and Montana educators reviewed test items and scoring guides for benchmarking. At that point, chief readers and selected QACs prepared scorer training materials.

Scoring staff from Measured Progress (including test developers) and Montana educators selected one or two anchor examples for each item score point. An additional six to ten responses per item were chosen as part of the training pack. The anchor pack consisted of midrange exemplars, while the training pack exemplars illustrated the range within each score point. The chief readers, who worked closely with QACs for each content area, facilitated the selection of response exemplars.

## SELECTING AND TRAINING SCORING STAFF

### QUALITY ASSURANCE COORDINATORS (QACS) AND VERIFIERS

Because the read-behinds performed by the QACs and verifiers moderated the scoring process and thus maintained the integrity of the scores, individuals chosen to fill those positions were selected for their accuracy. In addition, QACs, who train readers to score each item in their content areas, were selected for their ability to instruct and for their level of expertise in their content areas. For this reason, QACs typically are retired teachers who have demonstrated a high level of expertise in their respective disciplines. The ratio of QACs and verifiers to readers was approximately 1:11.

### TRAINING QUALITY ASSURANCE COORDINATORS AND VERIFIERS

To ensure that all QACs provided consistent training and feedback, the chief readers spent two days training and qualifying the QACs, and the QACs reviewed all items with the verifiers before scoring. In addition, QACs rotated among tables, supervising readers and reading behind verifiers, who in turn read behind a different table of readers each day.

### SELECTING READERS

Applicants were required to demonstrate their ability by participating in a preliminary scoring evaluation. The I-Score system enables Measured Progress to efficiently measure a prospective reader's ability to score student responses accurately. After participating in a training session, applicants were required to achieve at least 80% exact scoring agreement for a qualifying pack consisting of 20 responses to a predetermined item in their content area. Those 20 responses were randomly selected from a bank of approximately 150, all of which had been selected by QACs and approved by the chief readers and developers.

### TRAINING READERS

The QACs first applied the language of the scoring guide for an item to its anchor pack exemplars. Once discussion of the anchor pack had concluded, readers attempted to score the training pack exemplars correctly. The QACs then reviewed the training pack and answered any items readers had before actual scoring began. With this system, two aspects of scoring efficiency are in conflict. First, in order to minimize training expense, it is desirable to train each reader on as few items as possible. Second, to prevent reader drift and to minimize retraining requirements, it is desirable to score a given item in a brief period of time. But the lower the number of unique items each reader scores, the greater the number of readers required to score that item quickly. To minimize that conflict, we divided each subject area's readers into two or more groups. On the first day of scoring, each group was trained to score a different item. When a group had completed all of an item's responses, those readers were trained on another item (or set).

## SCORING ACTIVITIES

Student test booklets at grade level 4 and student response booklets at grade levels 8 and 10 were digitally scanned and scored on a file server for a dedicated, secure LAN. I-Score then distributed digital images of student responses to readers. Training and scoring took place over a period of approximately two weeks.

Items were randomly assigned to readers; thus, each item in a student's response booklet was more than likely scored by a different reader. By using the maximum possible number of readers for each student, the procedure effectively minimized error variance due to reader sampling. All common and matrix constructed-response items were scored once with a 2% read-behind to ensure consistency among readers and accuracy of individual readers.

## MONITORING READERS

After a reader scored a student response, I-Score determined whether that response should also be scored by another reader, scored by a QAC or verifier, or routed for special attention. To meet federal requirements, student responses indicating possible child abuse or suicidal tendencies were flagged by readers for OPI's attention ("crisis papers"). QACs and verifiers used I-Score to produce daily reader accuracy and speed reports. QACs and verifiers were able to obtain current reader accuracy reports and speed reports on-line at any time.

## GENERAL SCORING GUIDES

### TABLE 7-2: SHORT-ANSWER ITEMS

| Score Point | Description |
|---|---|
| 1 | The student’s response provides a complete and correct answer. |
| 0 | The student’s response is totally incorrect or too minimal to evaluate. |
| B | Blank/no response. |

### TABLE 7-3: CONSTRUCTED- RESPONSE ITEMS

| Score Point | Description |
|---|---|
| 4 | • The student completes all important components of the task and communicates ideas clearly.<br>• The student demonstrates in-depth understanding of the relevant concepts and/or processes.<br>• When instructed to do so, the student chooses more efficient and/or sophisticated processes.<br>• When instructed to do so, the student offers insightful interpretations or extensions (e.g., generalizations, applications, and analogies). |
| 3 | • The student completes the most important components of the task and communicates clearly.<br>• The student demonstrates understanding of major concepts even though he/she overlooks or misunderstands some less important ideas or details. |
| 2 | • The student completes most important components of the task and communicates those clearly.<br>• The student demonstrates that there are gaps in his/her conceptual understanding. |
| 1 | • The student shows minimal understanding.<br>• The student addresses only a small portion of the required task(s). |
| 0 | • The student’s response is totally incorrect or irrelevant. |
| B | • Blank/no response. |

# CHAPTER 8—ITEM ANALYSES

As noted in Brown (1983), “a test is only as good as the items it contains.” A complete evaluation of a test’s quality must include an evaluation of each item. Both the *Standards for Educational and Psychological Testing (1999)* and the *Code of Fair Testing Practices in Education (1988)* include standards for identifying quality items. Items should assess only knowledge or skills that are identified as part of the domain being tested and should avoid assessing irrelevant factors. They should also be unambiguous and free of grammatical errors, potentially insensitive content or language, and other confounding characteristics. Further, items must not unfairly disadvantage test takers from particular racial, ethnic, or gender groups.

Both qualitative and quantitative analyses are conducted to ensure that Montana CRT items meet these standards. Qualitative analyses are described in earlier sections of this report; this section focuses on the more quantitative evaluations. The statistical evaluations are presented in three parts: 1) difficulty indices, 2) item-test correlations, and 3) differential item functioning (DIF). The item analyses presented here are based on the statewide administration of the Montana CRT in spring 2005. About 10,315 grade 4 students, 11,720 grade 8 students, and 11,530 grade 10 students participated in the assessment.

## DIFFICULTY INDICES (P)

All multiple-choice, constructed-response and short-answer items were evaluated in terms of item difficulty according to standard classical test theory practices. Difficulty was defined as the average proportion of points achieved on an item, and was measured by obtaining the average score on an item and dividing by the maximum score for the item. Multiple-choice items were scored dichotomously (correct vs. incorrect), so for those items, the difficulty index is simply the proportion of students who correctly answered the item. The constructed-response items (five on each reading form and two on each math form) are scored polytomously, where a student can achieve a score of 0, 1, 2, 3, or 4; short-answer items (math computation) were scored 0 or 1. By computing the difficulty index as the average proportion of points achieved, the indices for the different item types are placed on a similar scale; the index ranges from 0.0 to 1.0 regardless of the item type. Although this index is traditionally described as a measure of difficulty, it is properly interpreted as an “easiness index” because larger values indicate easier items. An index of 0.0 indicates that all students received no credit for the item, and an index of 1.0 indicates that all students received full credit for the item.

Items that are answered correctly by almost all students provide little information about differences in student ability, but they do indicate knowledge or skills that have been mastered by most students. Similarly, items that are correctly answered by very few students may indicate knowledge or skills that have not yet been mastered by most students, but such items provide little information about differences in student ability. In general, to provide best measurement, difficulty indices should range from near-chance performance (.25 for four-option, multiple-choice items or essentially zero for constructed-response or short-answer items) to .90. Indices outside this range indicate items that were either too difficult or too easy for the target population. However, on a standards-referenced assessment such as the Montana CRT, it may be appropriate to include some items with very low or very high item difficulty values to ensure sufficient content coverage.

## ITEM-TEST CORRELATIONS (ITEM DISCRIMINATION)

A desirable feature of an item is that the higher-ability students perform better on the item than lower-ability students. The correlation between student performance on a single item and total test score is a commonly used measure of this characteristic of an item. Within classical test theory, the item-test correlation is referred to as the item's discrimination because it indicates the extent to which successful performance on an item discriminates between high and low scores on the test. For constructed-response items, the item discrimination index used was the Pearson product-moment correlation; for dichotomous items, the corresponding statistic is commonly referred to as a point-biserial correlation. The theoretical range of these statistics is –1 to +1, with a typical range from .2 to .6.

Discrimination indices can be thought of as measures of how closely an item assesses the same knowledge and skills assessed by other items contributing to the criterion total score. That is, the discrimination index can be thought of as a measure of construct consistency. In light of this interpretation, the selection of an appropriate criterion total score is crucial to the interpretation of the discrimination index. Because each form of the Montana CRT was constructed to be parallel in content, the criterion score selected for each item was the raw score total for each form. The analyses were conducted for each form separately.

## SUMMARY OF ITEM ANALYSIS RESULTS

Summary statistics of the difficulty and discrimination indices for each item are provided in Tables 8-1 through 8-3. Mean difficulty and discrimination indices, broken down by item type (multiple-choice/short-answer, constructed-response, and all items) are shown in Table 8-4 (standard deviations are shown in parentheses). In general, the item difficulty and discrimination indices are within generally acceptable and expected ranges. Very few items were answered correctly at near-chance or near-perfect rates. Similarly, the positive discrimination indices indicate that students who performed well on individual items tended to perform well overall. There were a small number of items with near-zero discrimination indices, but none were reliably negative. While it is not inappropriate to include items with low discrimination values or with very high or very low item difficulty values to ensure that content is appropriately covered, there were very few such cases on the Montana CRT.

A comparison of indices across grade levels is complicated because these indices are population dependent. Direct comparisons would require that either the items or students were common across groups. Since that is not the case, it can not be determined whether differences in performance across grade levels are due to differences in student ability or differences in item difficulty or both. However, one can say that for Reading, students in Grade 8 and 10 found their items more difficult than students in Grade 4 found their items.

Comparing the difficulty indices of multiple-choice and open-response (constructed-response or short-answer) items is inappropriate because multiple-choice items can be answered correctly by guessing. Thus, it is not surprising that the difficulty indices for multiple-choice items tend to be higher (indicating that students performed better on these items) than the difficulty indices for open-response items. Similarly, the partial credit allowed by constructed-response items is advantageous in the computation of item-test correlations, so the discrimination indices for these items tend to be larger than the discrimination indices of multiple-choice items.

The statistics in Tables 8-1 through 8-3 and those calculated for the full set of items in Table 8-4 are weighted according to the number of points contributed by each item. In the event that an item's statistics indicate it is flawed, the item is dropped from the operational form. An item may be dropped, for example, if more than one of the response options is a defensible answer, or if the item is misleading or unclear in some way. For the Montana CRT, one flawed item was excluded from the Grade 4 Math test and one flawed item was excluded from the Grade 10 Reading test. In addition, one grade 4 math item was excluded because it measured fractions, which are not included in Montana standards in grade 4.

### TABLE 8-1
### ITEM ANALYSIS: GRADE 4

| Content Area | | Difficulty | Discrimination |
|---|---|---|---|
| **Math** | Mean | 0.67 | 0.33 |
| | StDev | 0.17 | 0.08 |
| | Min | 0.27 | 0.16 |
| | Max | 0.93 | 0.52 |
| | Range | 0.66 | 0.36 |
| **Reading** | Mean | 0.72 | 0.40 |
| | StDev | 0.13 | 0.08 |
| | Min | 0.46 | 0.12 |
| | Max | 0.96 | 0.53 |
| | Range | 0.50 | 0.41 |

### TABLE 8-2
### ITEM ANALYSIS: GRADE 8

| Content Area | | Difficulty | Discrimination |
|---|---|---|---|
| **Math** | Mean | 0.47 | 0.30 |
| | StDev | 0.18 | 0.10 |
| | Min | 0.12 | 0.10 |
| | Max | 0.93 | 0.59 |
| | Range | 0.81 | 0.49 |
| **Reading** | Mean | 0.70 | 0.37 |
| | StDev | 0.14 | 0.08 |
| | Min | 0.39 | 0.15 |
| | Max | 0.93 | 0.55 |
| | Range | 0.54 | 0.40 |

### TABLE 8-3
### ITEM ANALYSIS: GRADE 10

| Content Area | | Difficulty | Discrimination |
|---|---|---|---|
| **Math** | Mean | 0.49 | 0.37 |
| | StDev | 0.15 | 0.11 |
| | Min | 0.14 | 0.15 |
| | Max | 0.78 | 0.70 |
| | Range | 0.64 | 0.55 |
| **Reading** | Mean | 0.69 | 0.35 |
| | StDev | 0.15 | 0.09 |
| | Min | 0.32 | 0.11 |
| | Max | 0.95 | 0.58 |
| | Range | 0.63 | 0.47 |

**TABLE 8-4**
**AVERAGE DIFFICULTY AND DISCRIMINATION OF DIFFERENT ITEM TYPES FOR EACH GRADE/CONTENT AREA COMBINATION**

|  |  |  | Item Type |  |  |
|---|---|---|---|---|---|
| Grade | Content Area |  | All | MC/SA | Constructed-Response |
| 4 | Reading | Difficulty | 0.72 (0.13) | 0.72 (0.12) | 0.49 (0.00) |
|  |  | Discrimination | 0.40 (0.08) | 0.39 (0.08) | 0.45 (0.02) |
|  |  | Number of Items | 54 | 52 | 2 |
|  | Mathematics | Difficulty | 0.67 (0.17) | 0.68 (0.16) | 0.59 (0.20) |
|  |  | Discrimination | 0.33 (0.08) | 0.32 (0.07) | 0.39 (0.11) |
|  |  | Number of Items | 58 | 53 | 5 |
| 8 | Reading | Difficulty | 0.70 (0.14) | 0.71 (0.14) | 0.52 (0.01) |
|  |  | Discrimination | 0.37 (0.08) | 0.36 (0.08) | 0.52 (0.04) |
|  |  | Number of Items | 54 | 52 | 2 |
|  | Mathematics | Difficulty | 0.47 (0.18) | 0.49 (0.17) | 0.24 (0.14) |
|  |  | Discrimination | 0.30 (0.10) | 0.28 (0.08) | 0.47 (0.10) |
|  |  | Number of Items | 60 | 55 | 5 |
| 10 | Reading | Difficulty | 0.69 (0.15) | 0.69 (0.15) | 0.53 (0.01) |
|  |  | Discrimination | 0.35 (0.09) | 0.35 (0.08) | 0.55 (0.04) |
|  |  | N | 58 | 56 | 2 |
|  | Mathematics | Difficulty | 0.49 (0.15) | 0.50 (0.15) | 0.36 (0.10) |
|  |  | Discrimination | 0.37 (0.11) | 0.35 (0.10) | 0.56 (0.10) |
|  |  | N | 65 | 60 | 5 |

*Note: Numbers shown in parentheses are standard deviations.

## DIFFERENTIAL ITEM FUNCTIONING (DIF)

The *Code of Fair Testing Practices in Education* explicitly states that subgroup differences in performance should be examined when sample sizes permit, and actions should be taken to make certain that differences in performance are due to construct-relevant, rather than irrelevant, factors. The *Standards for Educational and Psychological Testing* includes similar guidelines. As part of the effort to identify such problems, Montana CRT items were evaluated in terms of differential item functioning (DIF) statistics.

DIF procedures are designed to identify items for which subgroups of interest perform differently beyond the impact of differences in overall achievement. For the Montana CRT, the standardization DIF procedure (Dorans and Kulick, 1986) was employed to evaluate subgroup differences for three comparison groups: male/female, white/Native American, and white/Hispanic. This procedure calculates the difference in item performance for groups of students matched for achievement on the total test. That is, the average item performance is calculated for students at every total score, then an overall average is calculated weighting the total score distribution so it is the same for the two groups. The index ranges from –1 to 1 for multiple-choice items and is adjusted to the same scale for constructed-response items. Negative numbers indicate that the item was more difficult for female or non-white students. Dorans and Holland (1993) suggested that index values between –0.05 and 0.05 should be considered negligible. Most Montana CRT items fall within this range. Dorans and Holland further stated that items with values between –0.10 and –0.05 and between 0.05 and 0.10 (i.e., "low" DIF) should be inspected to ensure that no possible effect is overlooked, and that items with values outside the [–0.10, 0.10] range (i.e., "high" DIF) are more unusual and should be examined very carefully.

DIF indices indicate differential performance between two groups. That differential performance may or may not be indicative of bias in the test. Course-taking patterns, group differences in interests, or differences in school curricula can lead to DIF. If subgroup differences in performance are related to construct-relevant factors, the items should be considered for inclusion on a test.

Each item was categorized according to the guidelines adapted from Dorans and Holland (1993). Table 8-5 shows the number of items classified into each category separately by item type (multiple choice/short answer versus constructed response). Results are shown for male/female, white/Native American, and white/Hispanic comparisons. Table 8-6 provides the number of items in each of the three DIF categories that favor males or females, also separately by item type. There are some Montana CRT items categorized as "low" or "high" DIF. These indices must not be interpreted as indisputable evidence of bias. Both the *Code of Fair Testing Practices in Education* and the *Standards for Educational and Psychological Testing* assert that test items must be free from construct-irrelevant sources of differential difficulty. If subgroup differences in performance can be plausibly attributed to construct-relevant factors, the items may be included on a test. What is important is to determine if the cause of this differential performance is construct relevant.

For the Montana CRT, there were relatively few items flagged as having low or high DIF. The items that were flagged were reviewed for potential bias, and no obvious biases were detected. For this reason, and in order to ensure sufficient content coverage, no items were excluded from the test as a result of the DIF analyses.

### TABLE 8-5
### DIF ANALYSIS – ALL GRADES

| Grade | Content Area | Male/Female DIF Class | | | | | | | | | White/Native American DIF Class | | | | | | | | | White/Hispanic DIF Class | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | All | | | MC/SA | | | CR | | | All | | | MC/SA | | | CR | | | All | | | MC/SA | | | CR | | |
| | | A | B | C | A | B | C | A | B | C | A | B | C | A | B | C | A | B | C | A | B | C | A | B | C | A | B | C |
| 4 | Reading | 51 | 3 | 0 | 50 | 2 | 0 | 1 | 1 | 0 | 49 | 4 | 1 | 47 | 4 | 1 | 2 | 0 | 0 | 47 | 7 | 0 | 45 | 7 | 0 | 2 | 0 | 0 |
| | Math | 54 | 3 | 1 | 49 | 3 | 1 | 5 | 0 | 0 | 53 | 5 | 0 | 48 | 5 | 0 | 5 | 0 | 0 | 51 | 7 | 0 | 47 | 6 | 0 | 4 | 1 | 0 |
| 8 | Reading | 42 | 9 | 3 | 42 | 7 | 3 | 0 | 2 | 0 | 50 | 4 | 0 | 48 | 4 | 0 | 2 | 0 | 0 | 48 | 6 | 0 | 46 | 6 | 0 | 2 | 0 | 0 |
| | Math | 54 | 6 | 0 | 49 | 6 | 0 | 5 | 0 | 0 | 53 | 6 | 1 | 48 | 6 | 1 | 5 | 0 | 0 | 48 | 10 | 2 | 43 | 10 | 2 | 5 | 0 | 0 |
| 10 | Reading | 50 | 8 | 0 | 50 | 6 | 0 | 0 | 2 | 0 | 51 | 6 | 1 | 49 | 6 | 1 | 2 | 0 | 0 | 46 | 11 | 1 | 44 | 11 | 1 | 2 | 0 | 0 |
| | Math | 48 | 15 | 2 | 44 | 14 | 2 | 4 | 1 | 0 | 63 | 2 | 0 | 59 | 1 | 0 | 4 | 1 | 0 | 61 | 4 | 0 | 56 | 4 | 0 | 5 | 0 | 0 |

A = negligible DIF, B = low DIF, C = high DIF

### TABLE 8-6
### MALE VS. FEMALE DIFFERENTIAL ITEM FUNCTIONING (DIF) CATEGORIZATION BY ITEM TYPE (MULTIPLE-CHOICE/SHORT-ANSWER AND CONSTRUCTED-RESPONSE)

| Grade | Content Area | Item Type | Negligible DIF (A) | | | | Low DIF (B) | | | | High DIF (C) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Favor Female | Favor Male | N | % | Favor Female | Favor Male | N | % | Favor Female | Favor Male | N | % |
| 4 | Reading | MC/SA | 28 | 21 | 49 | 94 | 1 | 2 | 3 | 6 | 0 | 0 | 0 | 0 |
| | | CR | 1 | 0 | 1 | 50 | 1 | 0 | 1 | 50 | 0 | 0 | 0 | 0 |
| | Math | MC/SA | 27 | 22 | 49 | 92 | 1 | 2 | 3 | 6 | 0 | 1 | 1 | 2 |
| | | CR | 4 | 1 | 5 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | Reading | MC/SA | 25 | 17 | 42 | 81 | 3 | 4 | 7 | 13 | 0 | 3 | 3 | 6 |
| | | CR | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 100 | 0 | 0 | 0 | 0 |
| | Math | MC/SA | 24 | 25 | 49 | 89 | 1 | 5 | 6 | 11 | 0 | 0 | 0 | 0 |
| | | CR | 5 | 0 | 5 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | Reading | MC/SA | 32 | 18 | 50 | 89 | 1 | 5 | 6 | 11 | 0 | 0 | 0 | 0 |
| | | CR | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 100 | 0 | 0 | 0 | 0 |
| | Math | MC/SA | 29 | 17 | 46 | 77 | 5 | 7 | 12 | 20 | 0 | 2 | 2 | 3 |
| | | CR | 4 | 0 | 4 | 80 | 1 | 0 | 1 | 20 | 0 | 0 | 0 | 0 |

## ITEM RESPONSE THEORY ANALYSES

In addition to the classical test theory item analyses previously described, the Montana CRT tests were analyzed according to item response theory (IRT) models. IRT analyses were used, first, to place all 2005 forms on the same scale, and second, to equate the 2005 test to the previous year’s test. Details on the IRT calibration and equating procedures for the Montana CRT are provided in Chapter 10.

# CHAPTER 9—RELIABILITY

Although an individual item's performance is an important focus for evaluation, a complete evaluation of an assessment must also address the way items function together and complement one another. Tests that function well provide an accurate assessment of the student's level of ability. Unfortunately, no test can do this perfectly. A variety of factors can contribute to a given student's score being either higher or lower than his or her true ability. For example, a student may mis-read an item, or mistakenly fill in the wrong bubble when he or she knew the answer; similarly a student may get an item correct by guessing, even though he or she did not know the answer. Collectively, these extraneous factors that impact a student's score are referred to as measurement error. Any assessment includes some amount of measurement error; that is, no measurement can be perfectly accurate. This is true of academic assessments—no assessment can measure students perfectly accurately; some students will receive scores that underestimate their true ability, and other students will receive scores that overestimate their true ability. When tests have a high amount of measurement error student scores are very unstable. Students with high ability may get low scores or vice versa. Consequently, one cannot reliably tell a student's true level of ability with such a test. Assessments that have less measurement error (i.e., errors made are small on average and student scores on such a test will consistently represent their ability) are described as reliable.

There are a number of ways to estimate an assessment's reliability. One possible approach is to give the same test to the same students at two different points in time. If students receive the same scores on each test, then the extraneous factors affecting performance are small and the test is reliable. (This is referred to as test-retest reliability.) A potential problem with this approach is that students may remember items from the first administration or may have gained (or lost) knowledge or skills in the interim between the two administrations. A solution to the 'remembering items' problem is to give a different, but parallel test at the second administration. If student scores on each test correlate highly the test is considered reliable. (This is known as alternate forms reliability, because an alternate form of the test is used in each administration.) This approach, however, does not address the problem that students may have gained (or lost) knowledge or skills in the interim between the two administrations. In addition, the practical challenges of developing and administering parallel forms generally preclude the use of parallel forms reliability indices. One way to address these problems is to split the test in half and then correlate students' scores on the two half-tests; this in effect treats each half-test as a complete test. By doing this, the problems associated with an intervening time interval, and of creating and administering two parallel forms of the test, are alleviated. This is known as a split-half estimate of reliability. If the two half-test scores correlate highly, items on the two half-tests must be measuring very similar knowledge or skills. This is evidence that the items complement one another and function well as a group. This also suggests that measurement error will be minimal.

The split-half method requires a judgment regarding the selection of which items contribute to which half-test score. This decision may have an impact on the resulting correlation; different splits will give different estimates of reliability. Cronbach (1951) provided a statistic, α, that avoids this concern about the split-half method. Cronbach's α gives an estimate of the average of all possible splits for a given test. Cronbach's α is often referred to as a measure of internal consistency because it provides a measure of how well all the items in the test measure one single underlying ability. Cronbach's a is computed using the following formula:

$$a = \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} s^{2}\left(Y_i\right)}{s_x^{2}}\right]$$

where $i$ indexes the item
$n$ is the total number of items,
$s^{2}\left(Y_i\right)$ represents individual item variance, and
$s_x^{2}$ represents the total test variance

In addition to Cronbach's α, another approach to estimating the reliability for a test with differing item types (i.e., multiple-choice and constructed-response) is to assume that at least a small, but important, degree of unique variance is associated with item type (Feldt and Brennan, 1989). In contrast, Cronbach's coefficient α is built upon the assumption that there are no such local or clustered dependencies. A stratified version of coefficient α corrects for this problem by using the following formula:

$$a_{strat} = 1 - \frac{\sum_{j=1}^{k} s_{x_j}^{2}(1 - a_j)}{s_x^{2}}$$

where j indexes the subtests or categories,
$s_{x_i}^{2}$ represents the variance of each of the k individual subtests or categories,
$a_j$ is the unstratified Cronbach's $a$ coefficient for each subtest, and
$s_x^{2}$ represents the total test variance.

## RELIABILITY AND STANDARD ERRORS OF MEASUREMENT

Table 9-1 provides descriptive statistics, the overall Cronbach's α coefficient for each grade/content combination, and raw score standard errors of measurement. Tables 9-2 through 9-4 present Cronbach's α for each test form in each subject area (reading and mathematics), separately for each grade level. The tables also show reliability coefficients separately for multiple-choice/short-answer and constructed-response items, and stratified reliability coefficients that adjust for the fact that different item formats are included in the test.

Across the grades and content areas, the overall a coefficients, multiple-choice/short-answer a coefficients, and stratified a coefficients range from the mid-.80s to the low-.90s. There are little or no differences between the overall a and stratified a coefficients. The a coefficients for the constructed-response items are substantially lower, ranging from around 0.50 to around 0.75. These lower values can be explained, at least to some extent, by the fact that there are greater scoring inconsistencies for constructed-response items, as well as the relatively small numbers of these items on the test. Note that, for Reading, it is possible that the reliability coefficients are inflated as a result of passage-based item dependency.

### TABLE 9-1
### RELIABILITIES, STANDARD ERRORS OF MEASUREMENT, AND DESCRIPTIVE STATISTICS

| Grade | Content Area | N | Total Points | Mean | SD | Rel | SEM |
|---|---|---|---|---|---|---|---|
| **4** | **Reading** | 10314 | 60 | 41.59 | 10.28 | 0.91 | 3.05 |
| | **Mathematics** | 10311 | 64 | 42.00 | 10.11 | 0.88 | 3.47 |
| **8** | **Reading** | 11720 | 60 | 40.78 | 9.96 | 0.90 | 3.14 |
| | **Mathematics** | 11711 | 66 | 30.14 | 10.28 | 0.87 | 3.68 |
| **10** | **Reading** | 11529 | 64 | 43.01 | 10.47 | 0.90 | 3.30 |
| | **Mathematics** | 11505 | 71 | 33.47 | 13.53 | 0.92 | 3.85 |

## TABLE 9-2
## RELIABILITY ANALYSIS – GRADE 4

| Content Area | Reliability | Form | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Reading | **Coeff** a | 0.92 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.92 | 0.92 | 0.91 | 0.91 | 0.91 | 0.91 |
| | **MC/SA** a | 0.92 | 0.91 | 0.91 | 0.91 | 0.91 | 0.90 | 0.91 | 0.91 | 0.91 | 0.91 | 0.92 | 0.92 | 0.90 | 0.91 | 0.90 | 0.91 |
| | **CR** a | 0.59 | 0.50 | 0.57 | 0.52 | 0.52 | 0.54 | 0.53 | 0.52 | 0.57 | 0.47 | 0.47 | 0.53 | 0.46 | 0.54 | 0.56 | 0.54 |
| | **Strat** a | 0.92 | 0.92 | 0.91 | 0.91 | 0.92 | 0.91 | 0.92 | 0.91 | 0.91 | 0.91 | 0.92 | 0.92 | 0.91 | 0.91 | 0.91 | 0.91 |
| Mathematics | **Coeff** a | 0.89 | 0.89 | 0.88 | 0.88 | 0.88 | 0.87 | 0.88 | 0.88 | 0.88 | 0.88 | 0.89 | 0.89 | 0.87 | 0.89 | 0.87 | 0.88 |
| | **MC/SA** a | 0.88 | 0.88 | 0.87 | 0.87 | 0.87 | 0.86 | 0.87 | 0.87 | 0.87 | 0.87 | 0.88 | 0.88 | 0.86 | 0.88 | 0.86 | 0.87 |
| | **CR** a | 0.51 | 0.49 | 0.48 | 0.50 | 0.52 | 0.51 | 0.52 | 0.52 | 0.50 | 0.51 | 0.51 | 0.51 | 0.49 | 0.55 | 0.51 | 0.50 |
| | **Strat** a | 0.89 | 0.89 | 0.88 | 0.88 | 0.88 | 0.88 | 0.89 | 0.88 | 0.88 | 0.88 | 0.89 | 0.89 | 0.87 | 0.89 | 0.88 | 0.88 |

## TABLE 9-3
## RELIABILITY ANALYSIS – GRADE 8

| Content Area | Reliability | Form | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Reading | **Coeff** a | 0.91 | 0.90 | 0.90 | 0.90 | 0.91 | 0.90 | 0.89 | 0.90 | 0.90 | 0.89 | 0.91 | 0.90 | 0.91 | 0.90 | 0.90 | 0.89 |
| | **MC/SA** a | 0.90 | 0.90 | 0.89 | 0.89 | 0.90 | 0.89 | 0.89 | 0.89 | 0.89 | 0.89 | 0.90 | 0.90 | 0.90 | 0.89 | 0.90 | 0.89 |
| | **CR** a | 0.75 | 0.74 | 0.70 | 0.72 | 0.73 | 0.71 | 0.73 | 0.72 | 0.75 | 0.70 | 0.69 | 0.69 | 0.72 | 0.71 | 0.70 | 0.72 |
| | **Strat** a | 0.92 | 0.91 | 0.91 | 0.91 | 0.91 | 0.91 | 0.90 | 0.90 | 0.90 | 0.90 | 0.91 | 0.91 | 0.91 | 0.90 | 0.91 | 0.90 |
| Mathematics | **Coeff** a | 0.88 | 0.87 | 0.86 | 0.86 | 0.88 | 0.87 | 0.87 | 0.88 | 0.86 | 0.86 | 0.88 | 0.87 | 0.88 | 0.87 | 0.87 | 0.88 |
| | **MC/SA** a | 0.86 | 0.85 | 0.84 | 0.83 | 0.86 | 0.85 | 0.84 | 0.85 | 0.83 | 0.84 | 0.86 | 0.84 | 0.85 | 0.85 | 0.85 | 0.86 |
| | **CR** a | 0.62 | 0.55 | 0.57 | 0.55 | 0.60 | 0.58 | 0.59 | 0.59 | 0.59 | 0.55 | 0.62 | 0.60 | 0.57 | 0.56 | 0.58 | 0.60 |
| | **Strat** a | 0.88 | 0.87 | 0.87 | 0.86 | 0.88 | 0.87 | 0.87 | 0.88 | 0.86 | 0.86 | 0.88 | 0.87 | 0.88 | 0.88 | 0.87 | 0.88 |

## TABLE 9-4
## RELIABILITY ANALYSIS – GRADE 10

| Content Area | Reliability | Form | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Reading | **Coeff** a | 0.90 | 0.90 | 0.90 | 0.90 | 0.89 | 0.90 | 0.90 | 0.91 | 0.90 | 0.90 | 0.91 | 0.89 | 0.90 | 0.90 | 0.90 | 0.90 |
| | **MC/SA** a | 0.90 | 0.89 | 0.89 | 0.89 | 0.88 | 0.89 | 0.89 | 0.90 | 0.89 | 0.90 | 0.90 | 0.89 | 0.89 | 0.89 | 0.89 | 0.90 |
| | **CR** a | 0.71 | 0.64 | 0.66 | 0.66 | 0.67 | 0.68 | 0.67 | 0.73 | 0.65 | 0.70 | 0.70 | 0.67 | 0.67 | 0.68 | 0.68 | 0.68 |
| | **Strat** a | 0.91 | 0.91 | 0.91 | 0.91 | 0.90 | 0.91 | 0.90 | 0.91 | 0.90 | 0.91 | 0.91 | 0.90 | 0.91 | 0.90 | 0.91 | 0.91 |
| Mathematics | **Coeff** a | 0.92 | 0.92 | 0.91 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.91 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 |
| | **MC/SA** a | 0.90 | 0.90 | 0.90 | 0.91 | 0.91 | 0.90 | 0.91 | 0.91 | 0.91 | 0.90 | 0.90 | 0.91 | 0.90 | 0.90 | 0.91 | 0.91 |
| | **CR** a | 0.71 | 0.71 | 0.70 | 0.69 | 0.67 | 0.71 | 0.68 | 0.68 | 0.67 | 0.69 | 0.70 | 0.69 | 0.70 | 0.69 | 0.67 | 0.70 |
| | **Strat** a | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 | 0.93 | 0.92 | 0.92 | 0.92 | 0.92 |

## Reliability of Performance Level Categorization

All test scores contain measurement error; thus classifications based on test scores are also subject to measurement error. After the performance levels were specified and students were classified into those levels, empirical analyses were conducted to determine the statistical accuracy and consistency of the classifications.

## Accuracy

Accuracy refers to the extent to which decisions based on test scores match decisions that would have been made if the scores did not contain any measurement error. Accuracy must be estimated because errorless test scores do not exist.

## Consistency

Consistency measures the extent to which classification decisions based on test scores match the decisions based on scores from a second, parallel, form of the same test. Consistency can be evaluated directly from actual responses to test items if two complete, parallel, forms of the test are given to the same group of students. This is usually impractical, especially on lengthy tests. To overcome this issue, techniques have been developed to estimate both accuracy and consistency of classification decisions based on a single administration of a test. The technique developed by Livingston and Lewis (1995) was used for the Montana CRT because their technique can be used with both constructed-response and multiple-choice items.

## Calculating Accuracy

All of the accuracy and consistency estimation techniques described below make use of the concept of "true scores" in the sense of classical test theory. A true score is the score that would be obtained on a test that had no measurement error. It is a theoretical concept that cannot be observed, although it can be estimated. Following Livingston and Lewis (1995), the true-score distribution for the Montana CRT was estimated using a four-parameter beta distribution, which is a flexible model that allows for extreme degrees of skewness in test scores.

In the Livingston and Lewis method, the estimated "true scores" are used to classify students into their "true" performance category, which is labeled "true status." After various technical adjustments (which are described in Livingston and Lewis, 1995), a 4 × 4 contingency table was created for each content area test and grade level. The cells in the table are the proportion of students who were classified into each performance category by the actual (or observed) scores on the Montana CRT (i.e., observed status) and by the "true scores" (i.e., "true status").

## CALCULATING CONSISTENCY

To estimate consistency, the "true scores" are used to estimate the distribution of classifications on an independent, parallel test form. After statistical adjustments (see Livingston and Lewis, 1995), a new 4 × 4 contingency table was created for each test and grade level that shows the proportions of students who were classified into each performance category by the actual test and by another (hypothetical) parallel test form. Consistency, which is the proportion of students classified into exactly the same categories by the two forms of the test, is the sum of the diagonal for the new contingency table.

## KAPPA

Another way to measure consistency is to use Cohen's (1960) coefficient $\kappa$ (kappa), which assesses the proportion of consistent classifications after removing the proportion of consistent classification that would be expected by chance. Cohen's $\kappa$ can be used to estimate the classification consistency of a test from two parallel forms of the test. The second form in this case was the one estimated using the Livingston and Lewis (1995) method. Because $\kappa$ is corrected for chance, the values of $\kappa$ are lower than other consistency estimates.

## RESULTS OF ACCURACY, CONSISTENCY, AND KAPPA ANALYSES

Summaries of the Accuracy and Consistency analyses are provided in Tables 9-5 through 9-10. The first section of each table shows the overall accuracy and consistency indices as well as Kappa. The overall index is, as described above, the sum of the diagonal elements of the appropriate contingency table.

The second section of each table shows accuracy and consistency values, conditional upon performance level. In each case, the denominator is the number of students who were actually placed into a given performance level. For example, the conditional accuracy value is 0.7260 for the Proficient category for Grade 4 Math. This indicates that, of the students whose actual scores placed them in the Proficient category, 72.6% of them would be expected to be in the Proficient category if they were categorized according to their true score. Similarly, the corresponding consistency value of .6443 indicates that 64.43% of that same group of students would be expected to score in the Proficient category if a second, parallel test form were used.

For certain tests, concern may be greatest regarding decisions made about a particular threshold. For example, if a college gave credit to students who achieved an Advanced Placement test score of four or five, but not one, two, or three, one might be interested in the accuracy of the dichotomous decision, below four versus four or above. The third section of the summary tables shows information at each of the cut points. These values indicate the accuracy and consistency of the dichotomous decisions, either above or below the associated cut point. In addition, the false positive and false negative accuracy rates are also provided. These values are estimates of the proportion of students who were categorized above the cut when their true score would place them below the cut, and vice versa.

### TABLE 9-5
### ACCURACY AND CONSISTENCY -- GRADE 4 MATH

<table>
<tr><th colspan="7">Table 3.1.1.5. Accuracy and Consistency of Classification Indices</th></tr>
<tr><td rowspan="2"><b>Overall Indices</b></td><td colspan="2"><b>Accuracy</b></td><td colspan="2"><b>Consistency</b></td><td colspan="2"><b>Kappa (k)</b></td></tr>
<tr><td colspan="2">0.7304</td><td colspan="2">0.6363</td><td colspan="2">0.6272</td></tr>
<tr><td rowspan="5"><b>Indices Conditional on Level</b></td><td colspan="2"></td><td colspan="2"><b>Accuracy</b></td><td colspan="2"><b>Consistency</b></td></tr>
<tr><td colspan="2"><b><i>Novice</i></b></td><td colspan="2">0.8408</td><td colspan="2">0.7451</td></tr>
<tr><td colspan="2"><b><i>Nearing Proficiency</i></b></td><td colspan="2">0.5738</td><td colspan="2">0.4607</td></tr>
<tr><td colspan="2"><b><i>Proficient</i></b></td><td colspan="2">0.7260</td><td colspan="2">0.6443</td></tr>
<tr><td colspan="2"><b><i>Advanced</i></b></td><td colspan="2">0.8089</td><td colspan="2">0.6906</td></tr>
<tr><td rowspan="5"><b>Indices at Cut Points</b></td><td rowspan="2"></td><td colspan="4"><b>Accuracy</b></td><td rowspan="2"><b>Consistency</b></td></tr>
<tr><td><b>Accuracy</b></td><td><b><i>False Positives</i></b></td><td colspan="2"><b><i>False Negatives</i></b></td></tr>
<tr><td><b><i>N : NP</i></b></td><td>0.9248</td><td>0.0310</td><td colspan="2">0.0442</td><td>0.8941</td></tr>
<tr><td><b><i>NP : P</i></b></td><td>0.8921</td><td>0.0516</td><td colspan="2">0.0563</td><td>0.8491</td></tr>
<tr><td><b><i>P : A</i></b></td><td>0.9102</td><td>0.0547</td><td colspan="2">0.0351</td><td>0.8743</td></tr>
</table>

### TABLE 9-6
### ACCURACY AND CONSISTENCY -- GRADE 8 MATH

<table>
<tr><th colspan="7">Table 3.1.1.5. Accuracy and Consistency of Classification Indices</th></tr>
<tr><td rowspan="2"><b>Overall Indices</b></td><td colspan="2"><b>Accuracy</b></td><td colspan="2"><b>Consistency</b></td><td colspan="2"><b>Kappa (k)</b></td></tr>
<tr><td colspan="2">0.7795</td><td colspan="2">0.6943</td><td colspan="2">0.5560</td></tr>
<tr><td rowspan="5"><b>Indices Conditional on Level</b></td><td colspan="2"></td><td colspan="2"><b>Accuracy</b></td><td colspan="2"><b>Consistency</b></td></tr>
<tr><td colspan="2"><b><i>Novice</i></b></td><td colspan="2">0.7316</td><td colspan="2">0.4395</td></tr>
<tr><td colspan="2"><b><i>Nearing Proficiency</i></b></td><td colspan="2">0.6722</td><td colspan="2">0.4893</td></tr>
<tr><td colspan="2"><b><i>Proficient</i></b></td><td colspan="2">0.8103</td><td colspan="2">0.7676</td></tr>
<tr><td colspan="2"><b><i>Advanced</i></b></td><td colspan="2">0.7255</td><td colspan="2">0.5965</td></tr>
<tr><td rowspan="5"><b>Indices at Cut Points</b></td><td rowspan="2"></td><td colspan="4"><b>Accuracy</b></td><td rowspan="2"><b>Consistency</b></td></tr>
<tr><td><b>Accuracy</b></td><td><b><i>False Positives</i></b></td><td colspan="2"><b><i>False Negatives</i></b></td></tr>
<tr><td><b><i>N : NP</i></b></td><td>0.9920</td><td>0.0016</td><td colspan="2">0.0064</td><td>0.9880</td></tr>
<tr><td><b><i>NP : P</i></b></td><td>0.9464</td><td>0.0163</td><td colspan="2">0.0372</td><td>0.9227</td></tr>
<tr><td><b><i>P : A</i></b></td><td>0.8408</td><td>0.0916</td><td colspan="2">0.0676</td><td>0.7819</td></tr>
</table>

## TABLE 9-7
## ACCURACY AND CONSISTENCY -- GRADE 10 MATH

| **Table 3.1.1.5. Accuracy and Consistency of Classification Indices** | | | | | |
|---|---|---|---|---|---|
| **Overall Indices** | **Accuracy** | | **Consistency** | | **Kappa (κ)** |
| | 0.7991 | | 0.7192 | | 0.7154 |
| **Indices Conditional on Level** | | | **Accuracy** | | **Consistency** |
| | ***Novice*** | | 0.8749 | | 0.7955 |
| | ***Nearing Proficiency*** | | 0.7193 | | 0.6149 |
| | ***Proficient*** | | 0.8061 | | 0.7402 |
| | ***Advanced*** | | 0.8128 | | 0.7258 |
| **Indices at Cut Points** | | **Accuracy** | | | **Consistency** |
| | | **Accuracy** | ***False Positives*** | ***False Negatives*** | |
| | ***N : NP*** | 0.9549 | 0.0184 | 0.0266 | 0.9364 |
| | ***NP : P*** | 0.9269 | 0.0334 | 0.0396 | 0.8971 |
| | ***P : A*** | 0.9171 | 0.0443 | 0.0386 | 0.8838 |

## TABLE 9-8
## ACCURACY AND CONSISTENCY -- GRADE 4 READING

| **Table 3.1.1.5. Accuracy and Consistency of Classification Indices** | | | | | |
|---|---|---|---|---|---|
| **Overall Indices** | **Accuracy** | | **Consistency** | | **Kappa (κ)** |
| | 0.7945 | | 0.7124 | | 0.6988 |
| **Indices Conditional on Level** | | | **Accuracy** | | **Consistency** |
| | ***Novice*** | | 0.8511 | | 0.7560 |
| | ***Nearing Proficiency*** | | 0.6968 | | 0.5832 |
| | ***Proficient*** | | 0.7886 | | 0.7075 |
| | ***Advanced*** | | 0.8325 | | 0.7672 |
| **Indices at Cut Points** | | **Accuracy** | | | **Consistency** |
| | | **Accuracy** | ***False Positives*** | ***False Negatives*** | |
| | ***N : NP*** | 0.9668 | 0.0134 | 0.0198 | 0.9530 |
| | ***NP : P*** | 0.9375 | 0.0280 | 0.0345 | 0.9119 |
| | ***P : A*** | 0.8901 | 0.0539 | 0.0560 | 0.8453 |

## TABLE 9-9
## ACCURACY AND CONSISTENCY -- GRADE 8 READING

| **Table 3.1.1.5. Accuracy and Consistency of Classification Indices** | | | | | |
|---|---|---|---|---|---|
| **Overall Indices** | **Accuracy** | | **Consistency** | | **Kappa (κ)** |
| | 0.7683 | | 0.6856 | | 0.6820 |
| **Indices Conditional on Level** | | | **Accuracy** | | **Consistency** |
| | ***Novice*** | | 0.8649 | | 0.7868 |
| | ***Nearing Proficiency*** | | 0.5848 | | 0.4664 |
| | ***Proficient*** | | 0.7072 | | 0.6107 |
| | ***Advanced*** | | 0.8679 | | 0.8041 |
| **Indices at Cut Points** | | **Accuracy** | | | **Consistency** |
| | | **Accuracy** | ***False Positives*** | ***False Negatives*** | |
| | ***N : NP*** | 0.9417 | 0.0249 | 0.0334 | 0.9178 |
| | ***NP : P*** | 0.9179 | 0.0385 | 0.0436 | 0.8845 |
| | ***P : A*** | 0.9064 | 0.0503 | 0.0432 | 0.8689 |

## TABLE 9-10
## ACCURACY AND CONSISTENCY -- GRADE 10 READING

<table>
<tr><th colspan="6">Table 3.1.1.5. Accuracy and Consistency of Classification Indices</th></tr>
<tr><td rowspan="2"><b>Overall Indices</b></td><td colspan="2"><b>Accuracy</b></td><td colspan="2"><b>Consistency</b></td><td><b>Kappa (k)</b></td></tr>
<tr><td colspan="2">0.7636</td><td colspan="2">0.6783</td><td>0.6615</td></tr>
<tr><td rowspan="5"><b>Indices Conditional on Level</b></td><td colspan="2"></td><td colspan="2"><b>Accuracy</b></td><td><b>Consistency</b></td></tr>
<tr><td colspan="2"><b><i>Novice</i></b></td><td colspan="2">0.8418</td><td>0.7385</td></tr>
<tr><td colspan="2"><b><i>Nearing Proficiency</i></b></td><td colspan="2">0.5637</td><td>0.4424</td></tr>
<tr><td colspan="2"><b><i>Proficient</i></b></td><td colspan="2">0.7405</td><td>0.6568</td></tr>
<tr><td colspan="2"><b><i>Advanced</i></b></td><td colspan="2">0.8495</td><td>0.7786</td></tr>
<tr><td rowspan="5"><b>Indices at Cut Points</b></td><td rowspan="2"></td><td colspan="3"><b>Accuracy</b></td><td rowspan="2"><b>Consistency</b></td></tr>
<tr><td><b>Accuracy</b></td><td><b><i>False Positives</i></b></td><td><b><i>False Negatives</i></b></td></tr>
<tr><td><b><i>N : NP</i></b></td><td>0.9499</td><td>0.0198</td><td>0.0303</td><td>0.9290</td></tr>
<tr><td><b><i>NP : P</i></b></td><td>0.9192</td><td>0.0357</td><td>0.0451</td><td>0.8863</td></tr>
<tr><td><b><i>P : A</i></b></td><td>0.8920</td><td>0.0579</td><td>0.0500</td><td>0.8493</td></tr>
</table>

# CHAPTER 10— SCALING AND EQUATING

The purpose of equating is to ensure that scores obtained from different forms of a test are equivalent to each other. Equating may be used if multiple test forms are administered in the same year, as well as to equate one year's forms to those given in the previous year. Equating ensures that students are not given an unfair advantage or disadvantage because the test form given in one year is easier or harder than the form given in the other year. Once test scores for the forms are placed on an equivalent raw score scale, they then get translated, through the scaling process, to the score scale that is used for reporting.

## GENERAL RULES

- The goal is to have the entire common form be the equating set.
- Equating items cannot be changed from the version used in the previous form in any way.
- Whenever possible, items in the equating set should be selected so that they are within five positions of their location on the previous form.
- Passage sets selected for equating should consist of all, or most, of the items associated with the set.
- The equating set, as a whole, should mirror the characteristics of the common form in terms of content and statistics.

To determine the final set of equating items for each grade level and subject combination, a differential item functioning (DIF) approach using the delta method was applied. The 2005 and 2004 p-values of each multiple-choice item were transformed to the delta metric. The delta scale is an inverse normal transformation of percentage correct to a linear scale with a mean of 13 and standard deviation of 4 (Holland & Wainer, 1993). A high delta value indicates a difficult item. For constructed-response items, the average score divided by the maximum possible score, or adjusted p-value, was transformed to the delta metric. The delta values for the potential equating items were computed for each subject in each grade level.

Once all the delta values were calculated, a trend line was fit to the set of points. The perpendicular distance of each item to the regression line was then computed. Items that were not more than three standard deviations away from the regression line were used as equating items. As a result of the delta analyses, one item on the grade 4 math test was excluded for use as an equating item; all equating items were used for the remaining tests.

## IRT EQUATING

Equating for the Montana CRT used the *anchor-test-nonequivalent-groups* design described by Petersen, Kolen, & Hoover (1989). The fixed common-item IRT procedure was used, in which the anchor items from the previous year's administration were identified during this year's calibrations, and their IRT parameters were fixed to last year's values. This method results in all person and item parameters being on the same θ scale as last year. Because of the equating model that is used for the Montana CRT, the process of equating and scaling does not change the rank ordering of students, give more weight to particular items, or change students' performance-level classifications. Note that the groups of students who took the Montana CRT in 2003-04 and 2004-05 were not equivalent. Item Response Theory (IRT) is particularly useful in equating for nonequivalent groups (Allen & Yen, 1979).

IRT uses mathematical models to define a relationship between an unobserved measure of student ability, usually referred to as theta ($q$), and the probability (*p*) of getting a dichotomous item correct or of getting a particular score on a polytomous item. In IRT, it is assumed that all items are independent measures of the same construct or ability (i.e., the same $q$). There are several IRT models commonly used to specify the relationship between $q$ and *p*. For the Montana CRT tests, the 1 parameter logistic (1PL) model was used for multiple-choice and short-answer items and the partial credit model was used for the constructed-response items.

For polytomous items, the generalized partial credit model can be defined as:

$$P_{jk}(q) = \frac{\exp \sum_{v=0}^{k} \left[ Da_j \left( q - b_j + d_v \right) \right]}{\sum_{c=1}^{m} \exp \sum_{v=1}^{c} \left[ Da_j \left( q - b_j + d_v \right) \right]}$$

where *j* indexes the items,
*k* indexes students,
*a* represents item discrimination,
*b* represents item difficulty,
*d* represents category step parameter, and
*D* is a normalizing constant equal to 1.701.

In the case of the Montana CRT, the $a_j$ term in the above equation is equal to 1.0 for all items. For the dichotomous items, because there are no step parameters ($d_v$) the above equation reduces to the following:

$$P_j(q) = \frac{\exp\left(q - b_j\right)}{1 + \exp\left(q - b_j\right)}$$

For more information on IRT and IRT models the reader is referred to Hambleton and Swaminathan (1985).

The process of determining the specific mathematical relationship between $q$ and *p* is referred to as item calibration. Once items are calibrated, they are defined by a set of parameters which specify a non-linear relationship between $q$ and *p*. For more information about item calibration the reader is referred to Lord and Novick (1968) or Hambleton and Swaminathan (1985).

PARSCALE v3.5 (Muraki & Bock, 1999) was the software used to do the IRT analyses. The item parameter files resulting from the analyses are provided in Section V, Appendix B. Each item occupied only one block in the calibration run, and the 1.701 normalizing constant was used. A default convergence criterion of 0.001 was used, and all calibrations converged within 35 iterations.

## TRANSLATING RAW SCORES TO SCALED SCORES AND PERFORMANCE LEVELS

Montana CRT scores in each content area are reported on a scale that ranges from 200 to 300. Scaled scores supplement the Montana CRT performance-level results by providing information about the position of a student's results within a performance level. School- and district-level scaled scores are calculated by computing the average of student-level scaled scores. Students' raw scores, or total number of points, on the Montana CRT tests are translated to scaled scores using a data analysis process called **scaling**. Scaling simply converts raw points from one scale to another. In the same way that distance can be expressed in miles or kilometers, or monetary value can be expressed in terms of U.S. dollars or Canadian dollars, student scores on each Montana CRT could be expressed as raw scores (i.e., number right) or scaled scores. It is also important to notice that the raw score to scale score conversion formulae vary from CRT to CRT, analogous to how currency exchange formulae vary from country to country. For example, the scaling conversion formula for Montana's Grade 4 Reading CRT differs from that of the Grade 8 Reading CRT.

It is important to note that converting from raw scores to scaled scores does not change the students' performance-level classifications. Given the relative simplicity of raw scores, it is fair to ask why scaled scores are used in Montana CRT reports instead of raw scores. Foremost, scaled scores offer the advantage of simplifying the reporting of results across content areas, grade levels, and subsequent years. Because the standard-setting process typically results in different cut scores across content areas on a raw score basis, it is useful to transform these raw cut scores to a scale that is more easily interpretable and consistent. For the Montana CRT, a score of 225 is the cut score between the **Novice** and **Nearing Proficiency** performance levels. This is true regardless of which content area, grade, or year one may be concerned with. If one were to use raw scores, the raw cut score between **Novice** and **Nearing Proficiency** may be, for example, 35 in mathematics at grade 8, but may be 33 in mathematics at grade 10. Using scaled scores greatly simplifies the task of understanding how a student performed.

Cut points for the Montana CRT were originally set at the standard setting held in June, 2004. (See the 2004 Technical Manual for details on the standard setting meeting.) The original cut scores were established on the raw score metric. Therefore, in order to calculate scaling coefficients for the 2005 tests, it was first necessary to find the 2005 raw score equivalents to the 2004 cut points. The 2005 cut points were determined by first creating the test characteristic curves (TCCs) for both the 2004 and 2005 tests. From the 2004 TCC, the ?-scale equivalents of the 2004 cut points were calculated. These ?-scale cut points were then mapped through the 2005 TCC to find the 2005 cut points on the raw-score metric.

Once the 2005 raw score cut points were determined, the next step was to calculate the transformation coefficients that would be used to place students' raw scores onto the score scale used for reporting. As previously stated, student scores on the Montana CRT are reported in integer values from 200 to 300 with three scores representing cut scores on each assessment. Two of the three cut points (novice/nearing proficiency and nearing proficiency/proficient) were pre-set at 225 and 250, respectively; the third cut point, between proficient and advanced, was allowed to vary across tests, depending on where the raw score cuts were placed. Allowing the upper cut to float results in a single conversion equation for each test, which simplifies interpretation of scaled scores and their summary statistics. Table 10-1 presents the scaled score range for each performance level in each grade/content area combination.

TABLE 10-1

| Grade | Content Area | SCALED SCORE RANGE FOR EACH PERFORMANCE LEVEL | | | |
|---|---|---|---|---|---|
| | | Novice | Nearing Proficiency | Proficient | Advanced |
| 4 | Reading | 200–224 | 225–249 | 250–282 | 283–300 |
| | Mathematics | 200–224 | 225–249 | 250–285 | 286–300 |
| 8 | Reading | 200–224 | 225–249 | 250–282 | 283–300 |
| | Mathematics | 200–224 | 225–249 | 250–292 | 293–300 |
| 10 | Reading | 200–224 | 225–249 | 250–289 | 290–300 |
| | Mathematics | 200–224 | 225–249 | 250–287 | 288–300 |

The scaled scores are obtained by a simple linear transformation of the raw scores using the values of 225 and 250 on the scaled score metric and the associated 2005 raw score cut points to define the transformation. The scaling coefficients were calculated using the following formulae:

$$b = 225 - m(x_1)$$

$$m = \frac{225 - 250}{x_1 - x_2}$$

where m is the slope of the line providing the relationship between the raw and scaled scores, b is the intercept, $x_1$ is the cut score on the raw score metric for the novice/nearing proficiency cut, and $x_2$ is the cut score on the raw score metric for the nearing proficiency/proficient cut. Scaled scores were then calculated using the following linear transformation:

$$ss = m(x) + b$$

where x represents a student’s raw score. The values obtained using this formula were rounded to the nearest integer and truncated, as necessary, such that no student received a score below 200 or higher than 300.

# CHAPTER 11—REPORTING

The CRT assessments were designed to measure student performance against Montana's Content Standards. Consistent with this purpose, results on the CRT were reported in terms of performance levels that describe student performance in relation to these established state standards. There are four performance levels: *Advanced*, *Proficient*, *Nearing Proficiency*, and *Novice* (CRT Performance Level Descriptors, Scaled Score Ranges, and Raw Scores are described in greater detail in Appendix "E"). Students receive a separate performance-level classification (based on total scaled score) in each content area.

School- and system-level results are reported as the number and percentage of students attaining each performance level at each grade level tested. Disaggregations of students are also reported at the school and system levels. The CRT reports are

- Student Reports;
- Class Roster & Item-Level Reports;
- School Summary Reports; and
- System Summary Reports.

"Decision Rules" were formulated in late spring 2005 by OPI and Measured Progress to identify students, during the reporting process, to be excluded from school and system-level reports. A copy of these "Decision Rules" is included in this report as Appendix G.

State summary results were provided to OPI on confidential CDs and via a secure Web site. The report formats are included in Appendix F. These reports were shipped to System Test Coordinators on or before June 3, 2005 for distribution to schools within their respective systems/districts. System Test Coordinators and teachers were also provided with copies of the *Guide to Interpreting the 2005 Criterion-Referenced Test and CRT-Alternate Assessment Reports* and *iAnalyze,* to assist them in understanding the connection between the assessment and the classroom. The guide provides information about the assessment and the use of assessment results.

## IANALYZE

Using advanced Web technology, *iAnalyze* gives Montana educators and administrators the ability to filter data based on test year, grade level, and subject. Data can be sorted to isolate areas of strong or poor performance. Cross sections of data may be viewed by groupings based on demographics such as gender, Title 1 status, etc.

The confidential nature of the data therein necessitates the strict enforcement of site security. All transmissions are done over Secure Socket Layers (SSL). A system of user role definitions and permissions dictates the scope of access granted to individual users. Organizations (system or school levels) are given administrative power to grant or deny access to their data within the system, and have the ability to specify password durations, disable users, and create custom roles. Personnel using *iAnalyze* may be granted permission to view students' results at an organizational level, or only a select group as defined by the administrator. Each organization is also able to create custom data fields, and import/export functionality is provided. Predefined reports are included in the system, as is the ability to render and print additional copies.

### IANALYZE ENHANCEMENTS IN 2005

Below are a few of the features and enhancements added to the system for Montana in 2005:

- Maintaining history - Maintaining versions of information is another enhancement to *iAnalyze*. This feature allows the system to track changes to information, such as organization name or code, from year to year. Because of this, a name or identification change by an organization will not stop that organization from viewing or comparing historical data.
- Security enhancements - While simplifying user access to the system, we have strengthened security. Users will be allowed to log into the system with a user name and password that will be associated with the organization they belong to. *iAnalyze* uses SSL 128-bit encryption to ensure safe transmission of data to and from our servers.
- Charts and graphs - Charts and graphing capabilities have also been added to *iAnalyze*. Users will now be able to view data in a graphic manner rather than just as numbers.
- Import/Export - Importing and exporting features are enhanced. Users have more choices of how the data is imported into and exported from the system.
- "Assessment" and "Accountability" tabs were added to simplify the choice of requested data. "Assessment" selections contain raw data with a focus on school and classroom use. "Accountability" selections contain data in which "Decision Rules" or exclusions were applied. The focus user for this tab is school and system administrators.

# CHAPTER 12—VALIDITY SUMMARY

The purpose of this manual is to describe several technical aspects of the CRT in an effort to contribute to the accumulation of validity evidence to support CRT score interpretations. Because it is the interpretations of test scores that are evaluated for validity, not the test itself, this manual presents documentation to substantiate intended interpretations (AERA, 1999). Each of the chapters in this manual contributes important information to the validity argument by addressing one or more of the following aspects of the CRT: test development, test alignment, test administration, scoring, equating, item analyses, reliability, scaled scores, performance levels and reporting.

The CRT assessments are based on, and aligned to, Montana's Content Standards in Reading and Mathematics. Intended inferences from the CRT results are about student achievement on Montana's reading and mathematics content standards, and these achievement inferences are meant to be useful for program and instructional improvement and as a component of school accountability.

As stated in the overview chapter, the *Standards for Educational and Psychological Testing* (1999) provides a framework for describing sources of evidence that should be considered when constructing a validity argument. These sources include evidence based on the following five general areas: test content, response processes, internal structure, relationship to other variables, and consequences of testing. Although each of these sources may speak to a different *aspect* of validity, they are not distinct *types* of validity. Instead, each contributes to a body of evidence about the comprehensive validity of score interpretations.

A measure of test content validity is to determine how well the assessment tasks represent the curriculum and standards for each subject and grade level. This is informed by the item development process, including how the test blueprints and test items align to the curriculum and standards. Viewed through this lens provided by the Standards, evidence based on test content was extensively described in Chapters 2 through 5. Item alignment with Montana content standards; item bias, sensitivity and content appropriateness review processes; adherence to the test blueprint; use of multiple item types; use of standardized administration procedures, with accommodated options for participation; and appropriate test administration training are all components of validity evidence based on test content. As discussed earlier, all CRT test questions are aligned by Montana educators to specific Montana Content Standards, and undergo several rounds of review for content fidelity and appropriateness. Items are presented to students in multiple formats (constructed-response, short-answer and multiple-choice). Finally, tests are administered according to state-mandated standardized procedures, with allowable accommodations, and all test proctors are required to attend annual training sessions.

The scoring information in Chapter 7 describes the steps taken to train and monitor hand-scorers, as well as quality control procedures related to scanning and machine scoring. To speak to student response processes, however, additional studies would be helpful and might include an investigation of students' cognitive methods using think-aloud protocols.

Evidence based on internal structure is presented in great detail in the discussions of equating and item analyses in Chapters 8 and 9. Technical characteristics of the internal structure of the assessments are presented in terms of classical item statistics (item difficulty, item-test correlation), differential item functioning analyses, a variety of reliability coefficients, standard errors of measurement, and item response theory parameters and procedures. Each test is equated to the same grade and content test from the prior year in order to preserve the meaning of scores over time. In general, item difficulty and discrimination indices were in acceptable and expected ranges. Very few items were answered correctly at near-chance or near-perfect rates. Similarly, the positive discrimination indices indicate that most items were assessing consistent constructs, and students who performed well on individual items tended to perform well overall.

Evidence based on the consequences of testing is addressed in the scaled scores and reporting information in Chapters 10 and 11, as well as in the test interpretation guide, which is a separate document that is referenced in the discussion of reporting. Each of these chapters speaks to the efforts undertaken to promote accurate and clear information provided to the public regarding test scores. Scaled scores offer the advantage of simplifying the reporting of results across content areas, grade levels, and subsequent years. Performance levels provide users with reference points for mastery at each grade level, which is another useful and simple way to interpret scores. Several different standard reports are provided to stakeholders. In addition, a data analysis tool is provided to each school system to allow educators the flexibility to customize reports for local needs. Additional evidence of the consequences of testing could be supplemented with broader investigation of the impact of testing on student learning.

To further support the validity argument, additional studies to provide evidence regarding the relationship of CRT results to other variables include the extent to which scores from the CRT assessments converge with other measures of similar constructs, and the extent to which they diverge from measures of different constructs. Relationships among measures of the same or similar constructs can sharpen the meaning of scores and appropriate interpretations by refining the definition of the construct.

The evidence presented in this manual supports inferences of student achievement on the content represented on the Montana Content Standards for Reading and Mathematics for the purposes of program and instructional improvement and as a component of school accountability.

# SECTION IV—REFERENCES

Allen, Mary J. & Yen, Wendy M. (1979). Introduction to Measurement Theory. Belmont, CA: Wadsworth, Inc.

American Educational Research Association (AERA), American Psychological Association, and National Council on Measurement in Education 1999. *Standards for educational and psychological testing*. Washington, DC: AERA.

Bock, R. D., and E. Muraki. 1999. *PARSCALE: Parameter scaling of rating data* [Computer program]. Chicago, IL: Scientific Software.

Brown, F. G. 1983. *Principles of educational and psychological testing* 3rd ed. Fort Worth, TX: Holt, Rinehart, and Winston.

Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Dorans, N. J., and P. W. Holland. 1993. DIF detection and description. In P. W. Holland and H. Wainer (Eds.), *Differential item functioning* pp. 35–66. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dorans, N. J., and E. Kulick. 1986. Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, *23*, 355–368.

Hambleton, R. K., and W. J. van der Linden. 1997. *Handbook of modern item response theory*. New York: Springer-Verlag.

Hambleton, R. K., and H. Swaminathan. 1985. *Item response theory: Principles and applications*. Boston: Kluwer Academic Publishers.

Joint Committee on Testing Practices 1988. *Code of fair testing practices in education*. Washington, DC: National Council on Measurement in Education.

Lord, F.M., and M. R. Novick. 1968. *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989) Scaling, norming, and equating. In R.L. Linn (Ed.), Educational measurement (3rd ed., pp. 221-262).

# SECTION V—APPENDICES

**A. PTS Item and Bias Review Committees & Guidelines for PTS Reading Passage and Item Bias and Sensitivity Review**

**B. Item Parameter Files**

**C. Technical Advisory Committee**

**D. CRT Performance Level Descriptors, Scaled Scores, and Raw Scores**

**E. Report Shells**

**F. Reporting Decision Rules**

# Appendix A: PTS Item and Bias Review Committees and guidelines for PTS Reading Passage & Item Bias and Sensitivity Review

## National Bias Review Committee Members

| | First | Last | Organization | Bias Area |
|---|---|---|---|---|
| Ms. | Alada | Shinault-Small | Education Coordinator | African American |
| Dr. | Beverly | Chin | University of Montana | Asian American |
| Dr. | Barney | Berube | Maine Department of Education | ESL/Bilingual |
| Ms. | Sundra | Flansburg | Educational Development Center | Gender Equity |
| Mr. | David | Briseno | NM Association for Bilingual Ed. | Bilingual |
| Dr. | Roy | Howard | Western NM University | Native American |
| Ms. | Corri | Smith | Great Falls Public Schools | Native American |
| Ms. | Ellen | Honeyman | Worcester Public Schools - RETIRED | SPED |
| Ms. | Teri | Brogdon | University Center for Excellence | SPED/Gender |
| Ms. | Vaughn | Gross | Richardson, TX ISD - RETIRED | SPED |

## National Content Review Committee Members Reading

| Content | Grade | Title | | First | Last | Organization |
|---|---|---|---|---|---|---|
| Reading | 4 | Retired teacher | Ms. | Mariam | Miller | |
| Reading | 4 | Consultant | Dr. | Angelika | Pohl | |
| Reading | 4 | Elementary Principal | Ms. | Karen | Allen | Missoula County Public Schools |
| Reading | 4 | Teacher | Ms. | Becky | Sorenson | Missoula County Public Schools |
| Reading | 4 | Teacher | Ms. | Sarita | Kuhn | Gildford Colony School |
| Reading | 8 | Coordinator of English | Mr. | Robert | Zeeb | Newton Public Schools |
| Reading | 8 | Teacher | Mr. | Bruce | Fryar | Desonet School |
| Reading | 8 | Literacy Teacher | Ms. | Marilyn | Cron | Great Falls Public Schools |
| Reading | 8 | SPED Director | Ms. | Shannon | O'Brien | Dixon School |
| Reading | 10 | Writer/Editor | Ms. | Ursula | Szwast | |
| Reading | 10 | Title I Teacher/Rdg Specialist | Ms. | Vicky | Panasuk | Sidney Public Schools |
| Reading | 10 | Reading Specialist | Dr. | Janet | Hegedus | Big Sky High School |
| Reading | 10 | Teacher | Ms. | Marilyn | Beers | Hellgate High School |
| | | | | Jan | Katien | Measured Progress |
| | | | | Judy | Staten | Measured Progress |
| | | | | Ginny | Desmarais | Measured Progress |

## National Content Review Committee Members
## Mathematics

| Content | Grade | Title | | First | Last | Organization |
|---|---|---|---|---|---|---|
| Math | 4 | Math Teacher & Coach | Ms. | Jenny | Bland | Libby Public Schools |
| Math | 4 | Elementary educator | Ms. | Sheila | Murray | Fremont County School |
| Math | 4 | CEO | Ms. | Carol | Blunt-White | CBW Associates |
| Math | 8 | TIMSS Project Coordinator | Mr. | Steven | Chrostowski | Int'l Study Center at Boston College |
| Math | 8 | Teacher | Ms. | Vicki | Campbell | Missoula County Public Schools |
| Math | 8 | Associate Professor | Mr. | Ted | Hodgson | Montana State University |
| Math | 8 | Mathematics Teacher | Mr. | Lee | Brown | Hellgate High School |
| Math | 10 | Mathematics Teacher | Mr. | David | Bowie | Lewiston High School |
| Math | 10 | Asst. Sup of Curr & Instr. | Ms. | Cheryl | Wilson | Missoula County Public Schools |
| Math | 10 | Teacher | Ms. | Margaret | Aukshun | Billings West High School |
| Math | 10 | Deputy Executive Director | Dr. | Sharif | Shakrani | NAEP |
| | | | | Juliana | Cardone | Measured Progress |
| | | | | Sally | Schneider | Measured Progress |
| | | | | Alane | Fernald | Measured Progress |
| | | | | Christian | Citarella | Measured Progress |

# Guidelines
# for
# Progress Toward Standards
# Reading Passage & Item
# Bias and Sensitivity Review

Excerpted from a document by
Janice Dowd Scheuneman
Neal Kingston

(Updated by Rachel Slaughter according to suggestions of the
Progress Toward Standards National Bias Committee:
11/17/03)

## Bias, Sensitivity and Balance

- **Item Bias**

  Item bias stems from item context or content that is irrelevant to the curriculum elements being tested, but affects test scores of an identifiable subgroup of students. For example, several research studies have shown that if you couch a problem intended to test a student's ability to calculate percentages (curricularly relevant) in terms of batting averages (curricularly not relevant) girls will do less well relative to boys than if a non-sports context is used.

- **Sensitivity**

  Sensitivity concerns stem from issues that might offend or distract students, but that are not part of the curriculum framework being assessed. Affected students might be identifiable by race, ethnicity or sex, or by more subtle characteristics such as political leanings or religious beliefs.

  Sensitivity issues also include situations that might be disturbing to communities based on local events. For example, a reading item about teen suicide might affect the performance of test takers in a school where a student had recently taken his or her life. Sensitive issues are sometimes appropriate as part of instruction, but should be avoided in a test unless required to meet assessment specifications.

- **Curricular Context**

  Because both bias and sensitivity concerns must be considered in the context of the curriculum being measured, it is likely that some topics will be appropriate for some subject areas and inappropriate for others. For example, a item on evolution might be appropriate in a science test, or a item about suicide might be appropriate on a health test, but it might be inappropriate to have an item on evolution or health in a reading test.

- **Balance**

  Some bias and sensitivity issues arise at the level of test, not item. For example, it is not inappropriate to have a white male or a black female as the character in a item, but it would likely be inappropriate to have all characters in all items be white males or be black females.

## Goals for a Fair and Unbiased Examination

A fair and unbiased examination provides a context that permits all students to demonstrate their achievement and abilities. Students taking an unbiased test should feel that the test is appropriate for them. They should be able to feel that people like themselves are included as part of the assessment activity and are fairly represented in the examination materials. If this goal is to be met, examinations should:

- **Appropriately reflect the diversity of American society**

  Items, reading items, essay prompts, and illustrations should present boys and girls, men and women, including those from a variety of racial, ethnic, and language backgrounds, in a non-stereotypic manner.

  For example, the content of items should recognize differences of culture among citizens of the United States, e.g., hamburgers and French fries are not typical foods for all cultures in the country and some cultures adhere to a vegetarian diet.

Test contexts should be designed so that they are likely to be familiar to immigrants, newcomers and other groups new to the United States who may also have primary languages other than English. The context of the items should be those that are likely to be the most common possible across these diverse groups. For instance, American baseball and football may not be familiar to students across all cultural groups in the United States. If, however, the passage or item does not require prior knowledge but contains all the information the student needs to answer the item, unfamiliar contexts can be used.

The language of the items should also be reviewed to ensure that it is no more complex than is necessary to assess the knowledge or skill in order to be fair to students for whom English is not the primary language.

- **Use gender-fair language**

  If an item does not introduce an individual child or adult, it should be worded to be appropriate for either males or females.

- **Balance the representation of males and females**

  Each examination, or discrete part of an examination, should provide a balance of male and female figures.

- **Portray girls, women, and people of color in active roles**

  Women or girls and people of color should not be represented as passive recipients, observers of actions, or victims in need of rescue by others. If positions of power or status are suggested, the holders of these positions should be balanced in terms of gender with some representation of different racial or ethnic backgrounds.

  Portray contemporary women, girls and people of color as well as historical figures. Portray women, girls and people of color in ordinary, day-to-day situations, not just as historical figures or extraordinary individuals.

- **Show adults in nonstereotypic professions or work settings**

  Adults should be portrayed in ways that reflect the current reality of the workplace. Many positions once considered stereotypically male are now often held by women and people of color.

- **Distribute positions of power and status among members of different groups**

  Power and status should not be portrayed as the exclusive province of a single group. If such positions are used at all, they should belong to both men and women from a variety of backgrounds.

- **Recognize differences among family backgrounds**

  No single religious custom or family structure (such as mother, father, and children) should be represented as the norm. As appropriate to the item content, a variety of families should be reflected.

- **Acknowledge the contributions of women and people of color**

  To the extent that items identify the accomplishments of real people, examples should include women and people from various racial, ethnic, and language backgrounds.

- **Portray people with disabilities in a positive manner**

  When people with disabilities are portrayed, the material should emphasize their abilities and positive accomplishments rather than their disabilities.

## Fair Tests are Inclusive

Students will generally perform less well if they feel that the testing is for and about others. Presenting testing materials in ways that draw students in will help engage their attention and improve their motivation to perform well on the testing task.

- **All students should feel some connection with the test**

  Presentation of a variety of people from different racial/ethnic backgrounds increases the likelihood that students will see people like themselves in the testing situation.

- **Situations presented in the testing materials should be typical for most students or easy for them to imagine**

  Students should find it easy to identify with characters in stories or see themselves in the situations presented in test items. To the extent possible, settings should be familiar to all students.

- **Stereotypes serve to distance the student from the material**

  Students who are in groups presented in stereotypical situations are given the message that people see their group membership but not their individuality. It suggests that the test is more of the same material developed for somebody else.

- **Items with special appeal for boys or girls or for various racial/ethnic groups should be included**

  Research shows that students tend to perform better on items that are of special interest to them. Material should not be allowed to favor the interests of only one group.

- **Language used in items should be inclusive of both men/boys and women/girls**

  Avoid the generic he. Find an alternative such as recasting the item in the plural. Other approaches are suggested in the section, *Tips for Avoiding Generic Pronouns.*

# Stereotyping

### What is a stereotype?

- A standardized picture or mental image
- Oversimplified or exaggerated belief, uncritical judgment
- An unvarying pattern applied to all members of a group
- A lack of recognition of the individuality of the person
- Often accepted as fact
- Not always negative

### Stereotypes may be applied to many groups identified by

- gender
- race/ethnicity
- national origin
- religion
- language or language dialect
- political affiliation
- profession
- area of residence (inner city, rural, suburban)
- socioeconomic status
- age
- sexual orientation
- physical characteristics (blond, redhead, fat, short, tall)

## Dangers of Stereotypes

Stereotypes are not totally irrational and can be a convenient means of coping with diversity. On the other hand, even if they are positive, stereotypes can

- interfere with recognition of an individual’s qualities,
- reinforce preconceptions of people,
- eliminate the need to learn about individuals,
- insulate students from real person or group,

- affect judgments about people and how they are treated, and
- reinforce prejudice.
- contribute to hostility in relations between groups

Stereotypes may justify believing that

- a group is deserving of a particular fate;
- a group is dependent by nature and requires help from other groups (paternalism);
- a group is deficient or lacking in common human attributes such as emotional stability, honesty, industriousness, intelligence, leadership ability, morality, physical appearance, or physical capabilities;
- a group is deficient in qualities valued by society such as education, language proficiency, economic condition, political ideology, or professional status.

## Avoiding Stereotypes

Stereotypes can be well ingrained so that they sound natural and can be easy to miss, particularly those that do not seem negative or may even seem positive in tone.

One test of whether a statement about a racial/ethnic group or about a person from that group is acceptable is to substitute your own group or a member of your group for the one being discussed.

- Statements that seem neutral may be revealed as offensive.
- Statements that appear positive may come across as condescending or paternalistic.
- Statements with negative connotations should, of course, be avoided.

## Common Stereotypes

In order to assist in recognizing stereotypes, the following pages list some of the common stereotypes for major population subgroups: Women/girls, African Americans, Asian Americans, and Hispanic Americans. This list is not intended to be exhaustive, but only to illustrate some of the more common stereotypes that might be encountered.

Items that use the stereotypes in the following pages should be amended or deleted if possible.

## Stereotypes to Be Avoided: Girls/Women

- Overly concerned with physical appearance
- More concerned with home and family than career
- More intuitive, but less logical, than men or boys
- Physically less able than men

- Love to gossip and talk all the time
- All the same, regardless of race and ethnicity
- Spend large amounts of time and money shopping
- Disorganized and scatterbrained
- Emotional and cry easily, at the mercy of their hormones
- Emotions cloud judgment, making them unreliable decision makers
- Not team players
- Lack mechanical abilities and basic mathematics abilities
- Lack leadership qualities such as self-confidence, ambition, or assertiveness
- Less adequately prepared or less competent as professionals

### Stereotypes to Be Avoided: African American People

- Great athletes, physically powerful
- Musical, great sense of rhythm, terrific entertainers
- Speak “black” language
- Drive big cars and wear flashy clothes
- Loud, intense, have “attitudes”
- Don’t care about education
- Lazy and shiftless, don’t want to work
- Less adequately prepared or less competent as professionals
- Live in depressed urban areas
- Men often desert their families
- Children have children and become welfare mothers
- Less intelligent than other groups
- Use or sell drugs, carry weapons

## Stereotypes to Be Avoided: Hispanic American People

- Warm, expressive, and emotional
- Most often work in service or agricultural jobs
- Refuse to learn English
- Don’t value education
- Big on machismo, men dominate women
- Lazy and shiftless
- Not punctual and frequently procrastinate
- Don’t care if they’re on welfare
- Violent and hot tempered

## Stereotypes to Be Avoided: Asian American People

- Very intelligent, excellent scholars
- Hard working, ambitious, competitive
- Successful in business
- Strong family ties
- Quiet, polite, concerned with proper form
- Inscrutable, concerned with saving face
- Have marriages arranged between families
- Favor sons over daughters
- Prefer to live in ethnic neighborhoods
- Short, skinny, and wear glasses
- Predominantly refugees

## Stereotypes to be Avoided: Japanese American People

- Law-abiding
- Great imitators

- Sneaky
- Women are servile and obedient

**Stereotypes to be Avoided: Chinese American People**

- Great food
- Run good laundries and restaurants
- Love to gamble
- Use opium or its derivatives
- Cruel

## Gender Fair Language

Language that refers to people can be gender-neutral/gender-free or it can be gender-specific.

- **Gender-neutral language is inclusive**

  Gender-neutral language describes people with terms that can be used with either sex. This includes terms such as *student, teacher, writer, player, athlete, and parent.* In recent years, new terms have been introduced to refer to people who were once described by gender-specific terms that implied that job occupants were always men or always women. These new terms include *flight attendant, mail carrier, firefighter, police officer.*

- **Gender-specific terms should not be used to refer to people who may be either males or females**

  For many years, it was accepted practice to use masculine pronouns (*he, him, and his*) or the word *man,* as in *mankind,* generically to refer to either males or females. This is no longer the case. If the gender is nonspecific, gender-neutral terms should be used.

  Research shows that when *he* or *man* is used to refer to either sex, the majority of people perceive the reference as being to males only.

- **Some terms only appear to be gender-neutral**

  Terms such as *doctor, lawyer, politician, minister,* and *farmer* appear to be gender-neutral, but most people perceive them as men. In order to make these terms refer to women, special efforts may be needed by using a feminine pronoun or a name.

  *"Dr. Keesha Johnson treated both cats and dogs at her clinic."*

- **Gender-specific language may be appropriate in a gender fair test**

  Gender-specific terms such as *boy/girl, man/woman, mother/father* may promote gender-fair language in situations where use of a gender-neutral term may be interpreted in a gender-specific way.

**Examples:**

- *"Many children were accompanied by a parent when they had their vaccinations."*

  In this sentence many people might interpret *parent* to be *mother*. In such an instance, it would be better to say that

  *"Many children were accompanied by their mother or father when they had their vaccinations."*

- *"The players on the Spartans softball team traveled to their rival school by bus."*

  With the gender-neutral terms, many people might see the players on the team as boys. Again, the gender-specific terms may actually be more gender-fair.

  *"The boys and girls on the Spartans softball team traveled to their rival school by bus."*

- **Gender-fair language treats males and females equally**

  References to males and females should be symmetric with parallel terms used in the material: *Mr. Smith/Ms. Jones, John Smith/Janet Jones, man/woman, boy/girl, husband/wife.*

**Example:**

- *"Jorge and his sister each have nine stickers."*

  The girl in this sentence is defined only by her relationship to Jorge. She should at least be named, but probably for a test item her relationship with Jorge is not important and can be omitted.

  *"Jorge and Roselia each have nine stickers"*

- **Gender-fair language avoids unwarranted assumptions**

  Biased language often treats one type of person, family composition, or way of doing things as the norm, implying something deviant or substandard about those who do not conform.

## Balance and Equity

Both the individual items and the test, or test section as a whole, need to reflect equivalent treatment of different population subgroups.

- **Gender Balance**

  The number of references to males and to females should be nearly the same in subject-matter areas of the test.

- **Balance of Power**

  Some figures that are represented in the test items have more power or status than others. In most educational tests, the major power difference will be between child and adult, often a teacher.

  Adult figures should also include equal numbers of men and women as well as People of Color. If status differences exist among adult occupations represented, the higher status positions should be distributed among people from different groups.

- **Perspective**

  In attempting to convey a variety of environments in which students live, some situations will be more familiar to some students than to others. The items should not be overbalanced toward some settings, such as those that might be more familiar to middle-class families in the suburbs. No one family situation or environment should be presented as the norm.

- **Empowerment**

  Woman and People of Color should be portrayed as in control of their lives and destinies and independent of a need for more powerful groups to protect them and fight for their rights.

## Areas of Particular Interest for Girls and Boys

Because students tend to do better on materials that interest them, it is advisable to be aware of areas of particular interest.

Some areas that have been identified in research as having differential interest for girls and boys are as follows:

| Girls | Boys |
|---|---|
| Personal relations<br>Aesthetic, philosophic<br>Academic/school concerns<br>Home and family<br>Language, culture | Military, war, weapons<br>Sports<br>Physical Sciences<br>Mechanical, fixing/building things<br>Computers, computer games |

Selecting material that may appeal to different interests is appropriate and important. Items likely to be of greater interest to boys or to girls, however, should be balanced in each test form or in each module that will be used to make up a form.

### Interest and Prior Knowledge

Material that is interesting to examinees is likely to elicit greater attention to the material and increase motivation to read and understand the item being asked.

Greater interest can also lead to more experience and out-of-school learning about a topic. Care should be taken, therefore, to develop items that are not made easier for boys or for girls by prior knowledge or experience. If this occurs, the item may actually present an easier task for the group that is more interested in the topic. This, therefore, would be a biased item.

## Sensitive Material

An item that arouses strong emotions in students will most often be inappropriate in an educational achievement test. An emotional response may prevent them from clearly understanding the purpose of the item and the nature of the intended response. In addition, students who become upset during testing will become distracted from the task at hand and may fail to perform as well as they are able.

- **Personal Experience**

  If a child has had an experience like that described in an item, will the child be likely to find this upsetting?

  Examples: Death in the family, loss of a home

- **Privacy**

  Items should be avoided that may require students to reveal something about themselves or their families that they may not wish to discuss and feel is invasive.

- **Personal Values**

  Does the correct response depend on value judgments? This is particularly pertinent when considering how different racial or ethnic groups might respond.

- **Personal Reactions**

  Students should not be asked to discuss issues that they may find repugnant or discomforting. For example, students who oppose capital punishment may be distressed if asked to discuss only its merits.

**Avoid these topics:**

Child abuse or neglect
Incest
Rape
Abortion
Sex/Sexuality
Sexual orientation
Occult
Divorce
Parental conflict
Suicide

**Use these topics with caution:**

Death
Guns/Gun control
Homelessness
Animal rights
Racism/Sexism/Ageism
Religion
Family issues
Drugs/Alcohol/Tobacco
Murder
Pregnancy
Violence
Creation/Evolution

# Bibliography

American Psychological Association. (1994). Guidelines to Reduce Bias in Language. In *Publication Manual of the American Psychological Association,* Fourth Edition, pp. 46-60.

Interagency Committee for the Review of the Racial and Ethnic Standards. (July 1997). Recommendations to the Office of Management and Budget Concerning Changes to the Standards for the Classification of Federal Data on Race and Ethnicity. In *Federal Register Online* [wais.access.gpo.gov], Notices, Volume 62, Number 131, pages 36873-36946.

Maggio, R. (199-). *The Bias-Free Word Finder: A Dictionary of Nondiscriminatory Language.* Boston: Beacon Press.

National Evaluation Systems, Inc. (1991). *Bias Issues in Test Development.*

# Appendix B:  Item Parameter Files

## Grade 4 Math

## Grade 8 Math

## Grade 10 Math

## Grade 4 Reading

## Grade 8 Reading

## Grade 10 Reading

# APPENDIX B: ITEM PARAMETER FILES

## Grade 4 Math

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 189244 | 1 | 1 | -0.2122 | 0 | | | | |
| 189142 | 1 | 1 | -0.8656 | 0 | | | | |
| 189220 | 1 | 1 | -0.5228 | 0 | | | | |
| 189279 | 1 | 1 | -1.3334 | 0 | | | | |
| 189247 | 1 | 1 | 0.4159 | 0 | | | | |
| 189152 | 1 | 1 | -1.1155 | 0 | | | | |
| 189282 | 1 | 1 | -1.0664 | 0 | | | | |
| 189175 | 1 | 1 | 0.2417 | 0 | | | | |
| 189229 | 1 | 1 | 0.1115 | 0 | | | | |
| 189199 | 1 | 1 | -0.0079 | 0 | | | | |
| 189151 | 1 | 1 | -1.2395 | 0 | | | | |
| 206580 | 1 | 1 | -0.4828 | 0 | | | | |
| 189131 | 1 | 1 | -0.7958 | 0 | | | | |
| 189216 | 1 | 1 | -0.3327 | 0 | | | | |
| 189258 | 1 | 1 | -0.9582 | 0 | | | | |
| 189136 | 1 | 1 | -0.8868 | 0 | | | | |
| 206581 | 1 | 1 | 0.1904 | 0 | | | | |
| 214069 | 1 | 1 | 0.1899 | 0 | | | | |
| 189132 | 1 | 1 | -0.1681 | 0 | | | | |
| 189135 | 1 | 1 | 0.5045 | 0 | | | | |
| 189178 | 1 | 1 | -0.1816 | 0 | | | | |
| 189166 | 1 | 1 | -0.3923 | 0 | | | | |
| 166206 | 1 | 1 | -0.6299 | 0 | | | | |
| 165020 | 1 | 1 | -1.2011 | 0 | | | | |
| 165024 | 1 | 1 | 0.143 | 0 | | | | |
| 166218 | 1 | 1 | -0.5903 | 0 | | | | |
| 166289 | 1 | 1 | 0.0745 | 0 | | | | |
| 166364 | 1 | 1 | -1.0763 | 0 | | | | |
| 170345 | 1 | 1 | 0.2931 | 0 | | | | |
| 166366 | 1 | 1 | -0.1569 | 0 | | | | |
| 166291 | 1 | 1 | -0.1265 | 0 | | | | |
| 166403 | 1 | 1 | -0.094 | 0 | | | | |
| 189148 | 1 | 1 | -0.4769 | 0 | | | | |
| 189252 | 1 | 1 | 0.0855 | 0 | | | | |
| 189268 | 1 | 1 | -0.0722 | 0 | | | | |
| 189150 | 1 | 1 | 0.1937 | 0 | | | | |
| 189292 | 1 | 1 | -0.9251 | 0 | | | | |
| 189227 | 1 | 1 | -0.6842 | 0 | | | | |
| 214070 | 1 | 1 | 0.5493 | 0 | | | | |
| 189153 | 1 | 1 | -0.6578 | 0 | | | | |
| 206598 | 1 | 1 | 0.2294 | 0 | | | | |
| 189250 | 1 | 1 | 0.1591 | 0 | | | | |
| 189232 | 1 | 1 | -0.1808 | 0 | | | | |

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 206599 | 1 | 1 | -0.7425 | 0 | | | | |
| 189145 | 1 | 1 | -0.5727 | 0 | | | | |
| 189288 | 1 | 1 | -0.6858 | 0 | | | | |
| 189274 | 1 | 1 | -0.4327 | 0 | | | | |
| 189154 | 1 | 1 | -0.2159 | 0 | | | | |
| 189176 | 1 | 1 | -1.0035 | 0 | | | | |
| 189263 | 1 | 1 | -1.1417 | 0 | | | | |
| 189183 | 1 | 1 | -0.9104 | 0 | | | | |
| 206604 | 1 | 1 | 0.8647 | 0 | | | | |
| 214071 | 1 | 1 | -0.254 | 0 | | | | |
| 189297 | 1 | 1 | -0.683 | 0 | | | | |
| 189298 | 1 | 1 | 0.1922 | 0 | | | | |
| 189295 | 1 | 1 | -0.1998 | 0 | | | | |
| 189314 | 4 | 1 | -0.4132 | 0 | 0.5295 | -0.1527 | 0.3179 | -0.6947 |
| 206606 | 4 | 1 | 0.5705 | 0 | 0.5635 | 0.3911 | -0.8389 | -0.1157 |

## Grade 8 Math

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 189228 | 1 | 1 | -1.6277 | 0 | | | | |
| 189200 | 1 | 1 | -0.3744 | 0 | | | | |
| 210648 | 1 | 1 | 0.1509 | 0 | | | | |
| 214190 | 1 | 1 | -0.3706 | 0 | | | | |
| 189281 | 1 | 1 | -0.5879 | 0 | | | | |
| 210649 | 1 | 1 | 0.3973 | 0 | | | | |
| 189260 | 1 | 1 | 0.24 | 0 | | | | |
| 214191 | 1 | 1 | 0.3221 | 0 | | | | |
| 206726 | 1 | 1 | -0.3503 | 0 | | | | |
| 189221 | 1 | 1 | -0.3864 | 0 | | | | |
| 189302 | 1 | 1 | 0.3807 | 0 | | | | |
| 189233 | 1 | 1 | 0.15 | 0 | | | | |
| 210651 | 1 | 1 | 0.4889 | 0 | | | | |
| 189222 | 1 | 1 | -0.0264 | 0 | | | | |
| 214192 | 1 | 1 | 0.8287 | 0 | | | | |
| 189251 | 1 | 1 | 0.303 | 0 | | | | |
| 189259 | 1 | 1 | -0.1624 | 0 | | | | |
| 214194 | 1 | 1 | 0.4829 | 0 | | | | |
| 210654 | 1 | 1 | 0.6847 | 0 | | | | |
| 189273 | 1 | 1 | 0.3025 | 0 | | | | |
| 189196 | 1 | 1 | -0.0212 | 0 | | | | |
| 206722 | 1 | 1 | 0.2707 | 0 | | | | |
| 189210 | 1 | 1 | 0.301 | 0 | | | | |
| 206723 | 1 | 1 | 0.0019 | 0 | | | | |
| 165343 | 1 | 1 | -0.6342 | 0 | | | | |
| 210665 | 1 | 1 | -0.3791 | 0 | | | | |
| 165864 | 1 | 1 | 0.0661 | 0 | | | | |
| 165888 | 1 | 1 | -0.808 | 0 | | | | |
| 210669 | 1 | 1 | 0.6042 | 0 | | | | |
| 166520 | 1 | 1 | -0.3618 | 0 | | | | |
| 214196 | 1 | 1 | 0.2415 | 0 | | | | |
| 166330 | 1 | 1 | -0.8665 | 0 | | | | |
| 214204 | 1 | 1 | 0.0802 | 0 | | | | |
| 165802 | 1 | 1 | 0.24 | 0 | | | | |
| 189205 | 1 | 1 | -0.6615 | 0 | | | | |
| 189181 | 1 | 1 | -0.0585 | 0 | | | | |
| 210675 | 1 | 1 | -0.1074 | 0 | | | | |
| 214210 | 1 | 1 | 0.2664 | 0 | | | | |
| 206727 | 1 | 1 | 1.0065 | 0 | | | | |
| 214212 | 1 | 1 | -0.9586 | 0 | | | | |
| 210684 | 1 | 1 | 0.6797 | 0 | | | | |
| 189185 | 1 | 1 | 0.5051 | 0 | | | | |
| 206728 | 1 | 1 | -0.1876 | 0 | | | | |
| 217750 | 1 | 1 | 0.2136 | 0 | | | | |
| 210687 | 1 | 1 | 0.9259 | 0 | | | | |

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 210696 | 1 | 1 | 0.088 | 0 | | | | |
| 189243 | 1 | 1 | -0.0608 | 0 | | | | |
| 189209 | 1 | 1 | 0.0842 | 0 | | | | |
| 189304 | 1 | 1 | 0.4392 | 0 | | | | |
| 189187 | 1 | 1 | 0.2 | 0 | | | | |
| 189239 | 1 | 1 | -0.4422 | 0 | | | | |
| 210698 | 1 | 1 | 0.8964 | 0 | | | | |
| 189299 | 1 | 1 | -0.5456 | 0 | | | | |
| 189283 | 1 | 1 | 0.0258 | 0 | | | | |
| 189248 | 1 | 1 | 0.0317 | 0 | | | | |
| 189306 | 1 | 1 | 0.6136 | 0 | | | | |
| 189309 | 1 | 1 | 1.0103 | 0 | | | | |
| 189305 | 1 | 1 | 1.3779 | 0 | | | | |
| 206724 | 4 | 1 | 0.8724 | 0 | 0.3595 | -0.3025 | 0.1388 | -0.1958 |
| 189315 | 4 | 1 | 0.1367 | 0 | 0.563 | 0.5289 | -0.3491 | -0.7428 |

# Grade 10 Math

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 189333 | 1 | 1 | -0.4042 | 0 | | | | |
| 189359 | 1 | 1 | -0.2985 | 0 | | | | |
| 189371 | 1 | 1 | 0.5486 | 0 | | | | |
| 189386 | 1 | 1 | 0.0175 | 0 | | | | |
| 206690 | 1 | 1 | 0.3388 | 0 | | | | |
| 189324 | 1 | 1 | -0.3673 | 0 | | | | |
| 189369 | 1 | 1 | -0.2768 | 0 | | | | |
| 189368 | 1 | 1 | 0.3537 | 0 | | | | |
| 189358 | 1 | 1 | 0.4104 | 0 | | | | |
| 189362 | 1 | 1 | 0.0044 | 0 | | | | |
| 189334 | 1 | 1 | 0.5338 | 0 | | | | |
| 206691 | 1 | 1 | 1.0274 | 0 | | | | |
| 189328 | 1 | 1 | 0.3196 | 0 | | | | |
| 189352 | 1 | 1 | 0.248 | 0 | | | | |
| 189360 | 1 | 1 | 0.6926 | 0 | | | | |
| 189381 | 1 | 1 | -0.0054 | 0 | | | | |
| 214158 | 1 | 1 | -0.0246 | 0 | | | | |
| 189332 | 1 | 1 | -0.3012 | 0 | | | | |
| 189338 | 1 | 1 | 0.3058 | 0 | | | | |
| 189343 | 1 | 1 | -0.2258 | 0 | | | | |
| 206692 | 1 | 1 | -0.0505 | 0 | | | | |
| 189380 | 1 | 1 | 0.7657 | 0 | | | | |
| 206693 | 1 | 1 | -0.3721 | 0 | | | | |
| 189337 | 1 | 1 | -0.0135 | 0 | | | | |
| 166909 | 1 | 1 | -0.6483 | 0 | | | | |
| 166932 | 1 | 1 | 0.5151 | 0 | | | | |
| 166141 | 1 | 1 | -0.4552 | 0 | | | | |
| 166750 | 1 | 1 | -0.2848 | 0 | | | | |
| 166910 | 1 | 1 | 0.2447 | 0 | | | | |
| 166902 | 1 | 1 | 0.2587 | 0 | | | | |
| 166736 | 1 | 1 | 0.4658 | 0 | | | | |
| 166930 | 1 | 1 | -0.8887 | 0 | | | | |
| 166371 | 1 | 1 | 0.0132 | 0 | | | | |
| 166748 | 1 | 1 | 1.253 | 0 | | | | |
| 166941 | 1 | 1 | -0.4133 | 0 | | | | |
| 166130 | 1 | 1 | -0.4823 | 0 | | | | |
| 166966 | 1 | 1 | 0.0906 | 0 | | | | |
| 170231 | 1 | 1 | 0.607 | 0 | | | | |
| 166938 | 1 | 1 | -0.2088 | 0 | | | | |
| 189321 | 1 | 1 | -0.601 | 0 | | | | |
| 217405 | 1 | 1 | -0.4959 | 0 | | | | |
| 189348 | 1 | 1 | -0.4076 | 0 | | | | |
| 189376 | 1 | 1 | -0.769 | 0 | | | | |
| 189323 | 1 | 1 | 0.2109 | 0 | | | | |
| 206694 | 1 | 1 | -0.2794 | 0 | | | | |

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 189335 | 1 | 1 | -0.3445 | 0 | | | | |
| 189339 | 1 | 1 | -0.1607 | 0 | | | | |
| 189329 | 1 | 1 | 0.184 | 0 | | | | |
| 189387 | 1 | 1 | 0.5595 | 0 | | | | |
| 189366 | 1 | 1 | -0.2765 | 0 | | | | |
| 189356 | 1 | 1 | -0.113 | 0 | | | | |
| 189367 | 1 | 1 | -0.3035 | 0 | | | | |
| 189331 | 1 | 1 | 0.0704 | 0 | | | | |
| 189355 | 1 | 1 | 0.1615 | 0 | | | | |
| 189351 | 1 | 1 | 0.1858 | 0 | | | | |
| 189350 | 1 | 1 | -0.4751 | 0 | | | | |
| 206695 | 1 | 1 | 0.3287 | 0 | | | | |
| 189378 | 1 | 1 | 0.7426 | 0 | | | | |
| 189336 | 1 | 1 | -0.2019 | 0 | | | | |
| 189374 | 1 | 1 | -0.0636 | 0 | | | | |
| 189389 | 1 | 1 | 0.7148 | 0 | | | | |
| 189391 | 1 | 1 | -0.0913 | 0 | | | | |
| 189392 | 1 | 1 | 0.3517 | 0 | | | | |
| 189395 | 4 | 1 | 0.4135 | 0 | -0.7017 | 0.9562 | -0.2613 | 0.0068 |
| 189396 | 4 | 1 | 0.4584 | 0 | 0.2962 | 0.8389 | -0.6024 | -0.5328 |

# Grade 4 Reading

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 186862 | 1 | 1 | -1.9949 | 0 | | | | |
| 186864 | 1 | 1 | -1.6222 | 0 | | | | |
| 186866 | 1 | 1 | -0.7065 | 0 | | | | |
| 186868 | 1 | 1 | -0.4016 | 0 | | | | |
| 206797 | 1 | 1 | 0.1871 | 0 | | | | |
| 186946 | 1 | 1 | -0.9331 | 0 | | | | |
| 186956 | 1 | 1 | -0.6247 | 0 | | | | |
| 186948 | 1 | 1 | -0.5201 | 0 | | | | |
| 186949 | 1 | 1 | -1.0864 | 0 | | | | |
| 186951 | 1 | 1 | -0.6301 | 0 | | | | |
| 190994 | 1 | 1 | -0.271 | 0 | | | | |
| 181156 | 1 | 1 | -0.4092 | 0 | | | | |
| 190995 | 1 | 1 | -0.5913 | 0 | | | | |
| 181159 | 1 | 1 | -0.6425 | 0 | | | | |
| 181160 | 1 | 1 | -0.4092 | 0 | | | | |
| 191218 | 1 | 1 | 0.0327 | 0 | | | | |
| 191238 | 1 | 1 | -0.5417 | 0 | | | | |
| 181169 | 1 | 1 | -0.0158 | 0 | | | | |
| 181170 | 1 | 1 | -0.4648 | 0 | | | | |
| 190999 | 1 | 1 | -0.4829 | 0 | | | | |
| 206798 | 1 | 1 | 0.317 | 0 | | | | |
| 170978 | 1 | 1 | -1.1428 | 0 | | | | |
| 170975 | 1 | 1 | -0.5171 | 0 | | | | |
| 170976 | 1 | 1 | -1.6524 | 0 | | | | |
| 170973 | 1 | 1 | -0.812 | 0 | | | | |
| 214064 | 1 | 1 | -1.1603 | 0 | | | | |
| 171032 | 1 | 1 | -0.5447 | 0 | | | | |
| 171033 | 1 | 1 | -1.2186 | 0 | | | | |
| 171034 | 1 | 1 | -1.0471 | 0 | | | | |
| 171035 | 1 | 1 | -0.8588 | 0 | | | | |
| 171038 | 1 | 1 | -0.4255 | 0 | | | | |
| 181205 | 1 | 1 | -1.0111 | 0 | | | | |
| 214065 | 1 | 1 | -0.4123 | 0 | | | | |
| 199184 | 1 | 1 | -0.0502 | 0 | | | | |
| 206799 | 1 | 1 | -0.2939 | 0 | | | | |
| 181222 | 1 | 1 | -0.2272 | 0 | | | | |
| 214066 | 1 | 1 | -0.6776 | 0 | | | | |
| 181145 | 1 | 1 | -0.5622 | 0 | | | | |
| 181147 | 1 | 1 | -0.5671 | 0 | | | | |
| 206800 | 1 | 1 | -0.0831 | 0 | | | | |
| 190988 | 1 | 1 | -0.1778 | 0 | | | | |
| 181180 | 1 | 1 | -0.6592 | 0 | | | | |
| 214067 | 1 | 1 | -0.7259 | 0 | | | | |
| 181183 | 1 | 1 | -0.8608 | 0 | | | | |
| 190992 | 1 | 1 | -0.0903 | 0 | | | | |

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 182245 | 1 | 1 | -0.1172 | 0 | | | | |
| 181185 | 1 | 1 | -0.6377 | 0 | | | | |
| 190993 | 1 | 1 | -0.5299 | 0 | | | | |
| 181192 | 1 | 1 | 0.1029 | 0 | | | | |
| 181191 | 1 | 1 | 0.1627 | 0 | | | | |
| 181198 | 1 | 1 | -0.4512 | 0 | | | | |
| 181200 | 1 | 1 | -0.3901 | 0 | | | | |
| 192613 | 4 | 1 | 0.1573 | 0 | 1.8402 | 0.6064 | -0.9248 | -1.5219 |
| 192609 | 4 | 1 | 0.2213 | 0 | 0.7419 | 1.0613 | -0.6439 | -1.1592 |

# Grade 8 Reading

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 208994 | 1 | 1 | -0.1153 | 0 | | | | |
| 214181 | 1 | 1 | -0.4457 | 0 | | | | |
| 186464 | 1 | 1 | -0.816 | 0 | | | | |
| 214182 | 1 | 1 | -0.6471 | 0 | | | | |
| 214184 | 1 | 1 | 0.0912 | 0 | | | | |
| 186621 | 1 | 1 | -0.1449 | 0 | | | | |
| 186611 | 1 | 1 | -1.1827 | 0 | | | | |
| 186620 | 1 | 1 | -0.6159 | 0 | | | | |
| 186618 | 1 | 1 | -1.1724 | 0 | | | | |
| 186619 | 1 | 1 | -0.6136 | 0 | | | | |
| 186436 | 1 | 1 | -0.58 | 0 | | | | |
| 186441 | 1 | 1 | -0.3479 | 0 | | | | |
| 186439 | 1 | 1 | -0.4485 | 0 | | | | |
| 186445 | 1 | 1 | -0.5313 | 0 | | | | |
| 208996 | 1 | 1 | -0.063 | 0 | | | | |
| 214186 | 1 | 1 | -1.3541 | 0 | | | | |
| 186420 | 1 | 1 | -0.8784 | 0 | | | | |
| 186434 | 1 | 1 | -0.1936 | 0 | | | | |
| 208999 | 1 | 1 | -0.925 | 0 | | | | |
| 186414 | 1 | 1 | -0.196 | 0 | | | | |
| 208997 | 1 | 1 | 0.4383 | 0 | | | | |
| 214187 | 1 | 1 | -1.615 | 0 | | | | |
| 171173 | 1 | 1 | -1.3382 | 0 | | | | |
| 171177 | 1 | 1 | -1.2828 | 0 | | | | |
| 171179 | 1 | 1 | -0.919 | 0 | | | | |
| 171180 | 1 | 1 | -0.4816 | 0 | | | | |
| 214188 | 1 | 1 | -1.2834 | 0 | | | | |
| 171185 | 1 | 1 | 0.1377 | 0 | | | | |
| 171186 | 1 | 1 | -1.215 | 0 | | | | |
| 214189 | 1 | 1 | -0.5104 | 0 | | | | |
| 171193 | 1 | 1 | -1.0813 | 0 | | | | |
| 209004 | 1 | 1 | -0.5317 | 0 | | | | |
| 209003 | 1 | 1 | 0.1597 | 0 | | | | |
| 209005 | 1 | 1 | -0.446 | 0 | | | | |
| 209006 | 1 | 1 | -0.1262 | 0 | | | | |
| 186627 | 1 | 1 | -0.0103 | 0 | | | | |
| 186575 | 1 | 1 | -0.3707 | 0 | | | | |
| 209001 | 1 | 1 | -0.5403 | 0 | | | | |
| 209002 | 1 | 1 | -0.106 | 0 | | | | |
| 186589 | 1 | 1 | -0.5177 | 0 | | | | |
| 186583 | 1 | 1 | 0.17 | 0 | | | | |
| 186522 | 1 | 1 | -1.0356 | 0 | | | | |
| 186533 | 1 | 1 | -0.1638 | 0 | | | | |
| 209007 | 1 | 1 | -0.5347 | 0 | | | | |
| 186537 | 1 | 1 | -0.4627 | 0 | | | | |

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 186536 | 1 | 1 | 0.0457 | 0 | | | | |
| 186535 | 1 | 1 | -1.3307 | 0 | | | | |
| 186526 | 1 | 1 | -0.7726 | 0 | | | | |
| 186524 | 1 | 1 | -0.0339 | 0 | | | | |
| 186521 | 1 | 1 | -0.6936 | 0 | | | | |
| 186520 | 1 | 1 | 0.1176 | 0 | | | | |
| 186540 | 1 | 1 | -0.2453 | 0 | | | | |
| 186447 | 4 | 1 | -0.0035 | 0 | 1.5983 | 0.5135 | -0.7516 | -1.3602 |
| 186547 | 4 | 1 | 0.086 | 0 | 1.3638 | 0.6667 | -0.8555 | -1.175 |

# Grade 10 Reading

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 192143 | 1 | 1 | -0.9052 | 0 | | | | |
| 192145 | 1 | 1 | -1.1319 | 0 | | | | |
| 192148 | 1 | 1 | -0.2796 | 0 | | | | |
| 192147 | 1 | 1 | -0.9953 | 0 | | | | |
| 192149 | 1 | 1 | -1.0844 | 0 | | | | |
| 192155 | 1 | 1 | -0.907 | 0 | | | | |
| 192154 | 1 | 1 | -0.4557 | 0 | | | | |
| 192157 | 1 | 1 | -0.543 | 0 | | | | |
| 192158 | 1 | 1 | -0.7435 | 0 | | | | |
| 192160 | 1 | 1 | -0.0019 | 0 | | | | |
| 192167 | 1 | 1 | -0.0561 | 0 | | | | |
| 209008 | 1 | 1 | -0.2063 | 0 | | | | |
| 192179 | 1 | 1 | -0.5511 | 0 | | | | |
| 192174 | 1 | 1 | -1.2424 | 0 | | | | |
| 192170 | 1 | 1 | -0.0705 | 0 | | | | |
| 192169 | 1 | 1 | -0.3166 | 0 | | | | |
| 192171 | 1 | 1 | -0.6935 | 0 | | | | |
| 192180 | 1 | 1 | -0.9779 | 0 | | | | |
| 192176 | 1 | 1 | -0.5536 | 0 | | | | |
| 192172 | 1 | 1 | 0.0279 | 0 | | | | |
| 192175 | 1 | 1 | -1.1496 | 0 | | | | |
| 214109 | 1 | 1 | -0.4748 | 0 | | | | |
| 214110 | 1 | 1 | -1.3788 | 0 | | | | |
| 214111 | 1 | 1 | -0.5854 | 0 | | | | |
| 170771 | 1 | 1 | -1.8451 | 0 | | | | |
| 170773 | 1 | 1 | 0.258 | 0 | | | | |
| 170774 | 1 | 1 | -1.0665 | 0 | | | | |
| 214112 | 1 | 1 | 0.5547 | 0 | | | | |
| 170777 | 1 | 1 | -0.374 | 0 | | | | |
| 214120 | 1 | 1 | -0.802 | 0 | | | | |
| 170779 | 1 | 1 | -0.8991 | 0 | | | | |
| 170780 | 1 | 1 | -0.0917 | 0 | | | | |
| 170781 | 1 | 1 | -0.8995 | 0 | | | | |
| 214138 | 1 | 1 | -0.0873 | 0 | | | | |
| 170784 | 1 | 1 | 0.6614 | 0 | | | | |
| 192397 | 1 | 1 | -0.6373 | 0 | | | | |
| 192398 | 1 | 1 | -0.4117 | 0 | | | | |
| 192401 | 1 | 1 | 0.2424 | 0 | | | | |
| 192403 | 1 | 1 | 0.377 | 0 | | | | |
| 192404 | 1 | 1 | -0.7906 | 0 | | | | |
| 209011 | 1 | 1 | -0.5448 | 0 | | | | |
| 192422 | 1 | 1 | -0.5174 | 0 | | | | |
| 192421 | 1 | 1 | -0.9602 | 0 | | | | |
| 192426 | 1 | 1 | -0.7601 | 0 | | | | |
| 192419 | 1 | 1 | -1.199 | 0 | | | | |

| ITEM | MAX | A | B | C | D1 | D2 | D3 | D4 |
|---|---|---|---|---|---|---|---|---|
| 192204 | 1 | 1 | -0.2844 | 0 | | | | |
| 209012 | 1 | 1 | 0.2419 | 0 | | | | |
| 192196 | 1 | 1 | -0.9567 | 0 | | | | |
| 192185 | 1 | 1 | -0.8733 | 0 | | | | |
| 192192 | 1 | 1 | -0.0041 | 0 | | | | |
| 192199 | 1 | 1 | -0.5926 | 0 | | | | |
| 192183 | 1 | 1 | -0.2898 | 0 | | | | |
| 192187 | 1 | 1 | -0.7906 | 0 | | | | |
| 192194 | 1 | 1 | -0.3651 | 0 | | | | |
| 214144 | 1 | 1 | -0.4105 | 0 | | | | |
| 192190 | 1 | 1 | -0.1144 | 0 | | | | |
| 192181 | 4 | 1 | 0.2751 | 0 | 0.9495 | 0.6772 | -0.6391 | -0.9876 |
| 192209 | 4 | 1 | 0.0579 | 0 | 1.1145 | 0.9308 | -0.857 | -1.1883 |

# APPENDIX C: TECHNICAL ADVISORY COMMITTEE

| 2005 Technical Advisory Committee (TAC) Members | | | | |
|---|---|---|---|---|
| **First Name** | **Last Name** | **Position** | **Department** | **Organization** |
| Art | Bangert, Ph.D. | Assistant Professor | Adult and Higher Education | Montana State University |
| Rebecca | Walk, Ph.D. | Division Director | Special Education | Measured Progress |
| Liz | Burton, Ph.D. | Psychometrician | MDA | Measured Progress |
| Tim | Crockett | Vice President | Client Services | Measured Progress |
| Carolyn | Haug, Ph.D. | Asst. Division Director | Client Services | Measured Progress |
| Michael | Kozlow, Ph.D. | Program Director | Assessment Program | Northwest Regional Ed. Lab |
| Scott | Marion, Ph.D. | Vice-President | | Center for Assessment |
| Mike | Nering, Ph.D. | Psychometrician | MDA | Measured Progress |
| Madalyn | Quinlan | Chief Exectutive Officer | | OPI |
| Stanley | Rabinowitz, Ph.D. | Program Director | Assessment & Standards Development Services | WestEd |
| Nam | Raju, Ph. D. | Distinguished Professor | | Institute of Psychology |
| Steve | Sireci, Ph.D. | Associate Professor | | UMASS Amherst |
| Judy | Snow | State Assessment Director | | OPI |
| Wes | Snyder, Ph.D. | Assistant Vice Pres. | Research & Director of Office of International Programs | University of Montana |
| Kevin | Sweeney, Ph.D. | Division Director | MDA | Measured Progress |
| Bud | Williams | Asst. Superintendent | | OPI |

# APPENDIX D: CRT PERFORMANCE LEVEL DESCRIPTORS, SCALED SCORES AND RAW SCORES

## CRT Performance Level Descriptors

| | |
|---|---|
| **Advanced** | This level denotes superior performance. |
| **Proficient** | This level denotes solid academic performance for each benchmark. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter. |
| **Nearing Proficiency** | This level denotes that the student has partial mastery or prerequisite knowledge and skills fundamental for proficient work at each benchmark. |
| **Novice** | This level denotes that the student is beginning to attain the prerequisite knowledge and skills that are fundamental for work at each benchmark. |

## CRT Scaled Score Ranges for Performance Levels

**Grade 4**

| | Reading | Mathematics |
|---|---|---|
| Advanced | 283-300 | 286-300 |
| Proficient | 250-282 | 250-285 |
| Nearing Proficiency | 225-249 | 225-249 |
| Novice | 200-224 | 200-224 |

**Grade 8**

| | Reading | Mathematics |
|---|---|---|
| Advanced | 283-300 | 293-300 |
| Proficient | 250-282 | 250-292 |
| Nearing Proficiency | 225-249 | 225-249 |
| Novice | 200-224 | 200-224 |

**Grade 10**

| | Reading | Mathematics |
|---|---|---|
| Advanced | 290-300 | 288-300 |
| Proficient | 250-289 | 250-287 |
| Nearing Proficiency | 225-249 | 225-249 |
| Novice | 200-224 | 200-224 |

# CRT Cut Scores for Performance Levels

### TABLE D-1: CUT SCORES AND IMPACT DATA
### GRADE 4 READING

| Proficiency Level | Minimum Score | % in Level |
|---|---|---|
| Advanced | 48 | 30 |
| Proficient | 36 | 45 |
| Nearing Proficiency | 27 | 14 |
| Novice | -- | 11 |

### TABLE D-2: CUT SCORES AND IMPACT DATA
### GRADE 8 READING

| Proficiency Level | Minimum Score | % in Level |
|---|---|---|
| Advanced | 47 | 33 |
| Proficient | 39 | 31 |
| Nearing Proficiency | 33 | 16 |
| Novice | -- | 21 |

### TABLE D-3: CUT SCORES AND IMPACT DATA
### GRADE 10 READING

| Proficiency Level | Minimum Score | % in Level |
|---|---|---|
| Advanced | 50 | 31 |
| Proficient | 40 | 36 |
| Nearing Proficiency | 33 | 16 |
| Novice | -- | 16 |

### TABLE D-4: CUT SCORES AND IMPACT DATA
### GRADE 4 MATH

| Proficiency Level | Minimum Score | % in Level |
|---|---|---|
| Advanced | 52 | 18 |
| Proficient | 42 | 38 |
| Nearing Proficiency | 35 | 21 |
| Novice | -- | 23 |

### TABLE D-5: CUT SCORES AND IMPACT DATA
### GRADE 8 MATH

| Proficiency Level | Minimum Score | % in Level |
|---|---|---|
| Advanced | 41 | 16 |
| Proficient | 26 | 47 |
| Nearing Proficiency | 18 | 28 |
| Novice | -- | 9 |

### TABLE D-6: CUT SCORES AND IMPACT DATA
### GRADE 10 MATH

| Proficiency Level | Minimum Score | % in Level |
|---|---|---|
| Advanced | 46 | 21 |
| Proficient | 30 | 35 |
| Nearing Proficiency | 19 | 29 |
| Novice | -- | 15 |

**Raw-to-Scaled Score Correspondence**
Grade 4

| Raw Score | Reading Scaled Score | Mathematics Scaled Score |
|---|---|---|
| 0 | 200 | 200 |
| 1 | 200 | 200 |
| 2 | 200 | 200 |
| 3 | 200 | 200 |
| 4 | 200 | 200 |
| 5 | 200 | 200 |
| 6 | 200 | 200 |
| 7 | 200 | 200 |
| 8 | 200 | 200 |
| 9 | 200 | 200 |
| 10 | 200 | 200 |
| 11 | 200 | 200 |
| 12 | 200 | 200 |
| 13 | 200 | 200 |
| 14 | 200 | 200 |
| 15 | 200 | 200 |
| 16 | 200 | 200 |
| 17 | 201 | 200 |
| 18 | 203 | 200 |
| 19 | 206 | 200 |
| 20 | 209 | 200 |
| 21 | 211 | 200 |
| 22 | 214 | 200 |
| 23 | 217 | 200 |
| 24 | 219 | 200 |
| 25 | 222 | 200 |
| 26 | 224 | 200 |
| 27 | 227 | 200 |
| 28 | 230 | 202 |
| 29 | 232 | 205 |
| 30 | 235 | 209 |
| 31 | 238 | 213 |
| 32 | 240 | 216 |
| 33 | 243 | 220 |
| 34 | 246 | 223 |
| 35 | 248 | 227 |
| 36 | 251 | 231 |
| 37 | 254 | 234 |
| 38 | 256 | 238 |
| 39 | 259 | 242 |
| 40 | 262 | 245 |
| 41 | 264 | 249 |
| 42 | 267 | 252 |
| 43 | 269 | 256 |
| 44 | 272 | 260 |
| 45 | 275 | 263 |
| 46 | 277 | 267 |

|  |  |  |
|---|---|---|
| 47 | 280 | 270 |
| 48 | 282 | 274 |
| 49 | 285 | 278 |
| 50 | 288 | 281 |
| 51 | 291 | 285 |
| 52 | 293 | 288 |
| 53 | 296 | 292 |
| 54 | 299 | 296 |
| 55 | 300 | 299 |
| 56 | 300 | 300 |
| 57 | 300 | 300 |
| 58 | 300 | 300 |
| 59 | 300 | 300 |
| 60 | 300 | 300 |
| 61 |  | 300 |
| 62 |  | 300 |
| 63 |  | 300 |
| 64 |  | 300 |

**Raw-to-Scaled Score Correspondence**
Grade 8

| Raw Score | Reading Scaled Score | Mathematics Scaled Score |
|---|---|---|
| 0 | 200 | 200 |
| 1 | 200 | 200 |
| 2 | 200 | 200 |
| 3 | 200 | 200 |
| 4 | 200 | 200 |
| 5 | 200 | 200 |
| 6 | 200 | 200 |
| 7 | 200 | 200 |
| 8 | 200 | 200 |
| 9 | 200 | 202 |
| 10 | 200 | 205 |
| 11 | 200 | 208 |
| 12 | 200 | 211 |
| 13 | 200 | 213 |
| 14 | 200 | 216 |
| 15 | 200 | 219 |
| 16 | 200 | 222 |
| 17 | 200 | 224 |
| 18 | 200 | 228 |
| 19 | 200 | 231 |
| 20 | 200 | 233 |
| 21 | 200 | 236 |
| 22 | 200 | 239 |
| 23 | 200 | 242 |
| 24 | 200 | 245 |
| 25 | 200 | 248 |
| 26 | 201 | 251 |
| 27 | 205 | 253 |
| 28 | 209 | 256 |
| 29 | 213 | 259 |
| 30 | 217 | 262 |
| 31 | 221 | 265 |
| 32 | 224 | 268 |
| 33 | 229 | 271 |
| 34 | 233 | 274 |
| 35 | 237 | 276 |
| 36 | 241 | 279 |
| 37 | 245 | 282 |
| 38 | 249 | 285 |
| 39 | 253 | 288 |
| 40 | 257 | 291 |
| 41 | 261 | 294 |
| 42 | 265 | 296 |
| 43 | 269 | 299 |

| | | |
|---|---|---|
| 44 | 273 | 300 |
| 45 | 277 | 300 |
| 46 | 281 | 300 |
| 47 | 285 | 300 |
| 48 | 289 | 300 |
| 49 | 293 | 300 |
| 50 | 297 | 300 |
| 51 | 300 | 300 |
| 52 | 300 | 300 |
| 53 | 300 | 300 |
| 54 | 300 | 300 |
| 55 | 300 | 300 |
| 56 | 300 | 300 |
| 57 | 300 | 300 |
| 58 | 300 | 300 |
| 59 | 300 | 300 |
| 60 | 300 | 300 |
| 61 | | 300 |
| 62 | | 300 |
| 63 | | 300 |
| 64 | | 300 |
| 65 | | 300 |
| 66 | | 300 |

**Raw-to-Scaled Score Correspondence**
Grade 10

| Raw Score | Reading Scaled Score | Mathematics Scaled Score |
|---|---|---|
| 0 | 200 | 200 |
| 1 | 200 | 200 |
| 2 | 200 | 200 |
| 3 | 200 | 200 |
| 4 | 200 | 200 |
| 5 | 200 | 200 |
| 6 | 200 | 200 |
| 7 | 200 | 200 |
| 8 | 200 | 202 |
| 9 | 200 | 204 |
| 10 | 200 | 206 |
| 11 | 200 | 209 |
| 12 | 200 | 211 |
| 13 | 200 | 213 |
| 14 | 200 | 215 |
| 15 | 200 | 218 |
| 16 | 200 | 220 |
| 17 | 200 | 222 |
| 18 | 200 | 224 |
| 19 | 200 | 227 |
| 20 | 200 | 229 |
| 21 | 200 | 231 |
| 22 | 200 | 233 |
| 23 | 200 | 236 |
| 24 | 200 | 238 |
| 25 | 200 | 240 |
| 26 | 200 | 242 |
| 27 | 203 | 245 |
| 28 | 206 | 247 |
| 29 | 210 | 249 |
| 30 | 214 | 252 |
| 31 | 218 | 254 |
| 32 | 222 | 256 |
| 33 | 226 | 258 |
| 34 | 230 | 261 |
| 35 | 234 | 263 |
| 36 | 237 | 265 |
| 37 | 241 | 267 |
| 38 | 245 | 270 |
| 39 | 249 | 272 |
| 40 | 253 | 274 |
| 41 | 257 | 276 |
| 42 | 261 | 279 |
| 43 | 265 | 281 |

| | | |
|---|---|---|
| 44 | 268 | 283 |
| 45 | 272 | 285 |
| 46 | 276 | 288 |
| 47 | 280 | 290 |
| 48 | 284 | 292 |
| 49 | 288 | 295 |
| 50 | 292 | 297 |
| 51 | 296 | 299 |
| 52 | 300 | 300 |
| 53 | 300 | 300 |
| 54 | 300 | 300 |
| 55 | 300 | 300 |
| 56 | 300 | 300 |
| 57 | 300 | 300 |
| 58 | 300 | 300 |
| 59 | 300 | 300 |
| 60 | 300 | 300 |
| 61 | 300 | 300 |
| 62 | 300 | 300 |
| 63 | 300 | 300 |
| 64 | 300 | 300 |
| 65 | | 300 |
| 66 | | 300 |
| 67 | | 300 |
| 68 | | 300 |
| 69 | | 300 |
| 70 | | 300 |
| 71 | | 300 |

# APPENDIX E: REPORT SHELLS

## Student Report

## Class Roster & Item-Level Report

## School Summary Report

## System Summary Report

# CRT Performance Level Descriptors

The Performance Level Descriptors below describe students' knowledge, skills, and abilities in a content area. These descriptions provide a picture or profile of student achievement at the four performance levels: **Advanced**, **Proficient**, **Nearing Proficiency**, and **Novice**.

**Advanced**
This level denotes superior performance.

**Proficient**
This level denotes solid academic performance for each benchmark. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.

**Nearing Proficiency**
This level denotes that the student has partial mastery or prerequisite knowledge and skills fundamental for proficient work at each benchmark.

**Novice**
This level denotes that the student is beginning to attain the prerequisite knowledge and skills that are fundamental for work at each benchmark.

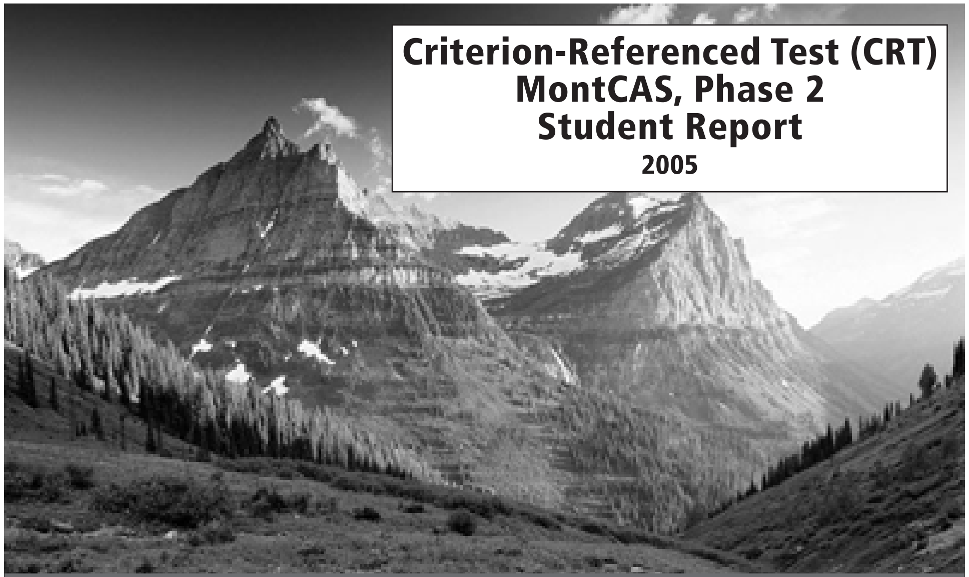
**Score Ranges**

| | Reading | Math |
|---|---|---|
| Advanced | (283-300) | (286-300) |
| Proficient | (250-282) | (250-285) |
| Nearing Proficiency | (225-249) | (225-249) |
| Novice | (200-224) | (200-224) |

**For more information regarding student assessments in Montana, check out the Office of Public Instruction's Parents Page at http://www.opi.mt.us/parents/.**

**OPI Contact**
**Judy Snow, State Assessment Director**
**406-444-3656**
**jsnow@mt.gov**

OPI
**Linda McCulloch, Superintendent**
**Montana Office of Public Instruction**
**PO Box 202501**
**Helena, Montana 59620-2501**
**http://www.opi.mt.gov**



# Criterion-Referenced Test (CRT)
# MontCAS, Phase 2
# Student Report
**2005**

Student Name:
School:
System:
Grade: 04

Dear Parents/Guardians:

This report contains the results of the second year of the Montana Comprehensive Assessment System Criterion-Referenced Test (CRT) that your child took in March. The major purpose of the CRT is to provide schools with solid information to evaluate and improve curriculum and instruction to help all students meet Montana's reading and mathematics standards. This report provides important information about your child's performance on the assessment, along with state results.

The CRT contains multiple-choice and short-answer questions. The test measures a student's knowledge of subject matter identified in the Montana State Standards for Reading and Mathematics. Your child's results in reading and mathematics are reported in one of four performance levels. These performance levels are defined on the back cover of this report.

It is important to remember that the CRT is just one measure of your child's academic progress. Your local school staff can provide further information about your child's performance in school. The CRT, which is required by the No Child Left Behind Act, is part of an ongoing statewide educational improvement process. Working together, we can ensure that Montana's children continue to receive a high-quality education.

Sincerely,

Linda McCulloch
Montana Superintendent of Public Instruction

# How did YOUR CHILD do on the CRT?

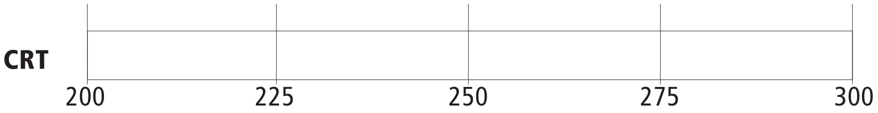
## Scaled Scores on the CRT

The criterion-referenced test (CRT) is designed to measure student performance against the learning goals described in the Montana Content Standards (http://www.opi.state.mt.us/standards/index.html). Consistent with this purpose, results on the CRT are reported according to performance levels that describe student performance in relation to the established state standards. There are four performance levels: **Advanced**, **Proficient**, **Nearing Proficiency**, and **Novice**. Your child's performance levels in reading and mathematics are based on a total scaled score in each content area. Scaled scores in each content area range from 200 to 300. Your child's performance levels, based on the scaled scores, are shown in the bar graphs below.

## Scaled Scores

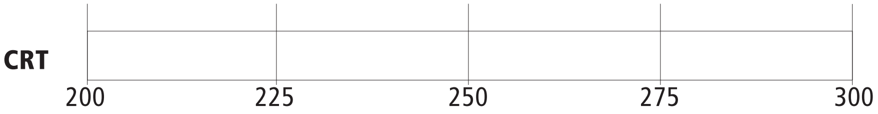
### STUDENT RESULTS FOR READING

**Performance Level:**
**Student Scaled Score:**



### STUDENT RESULTS FOR MATHEMATICS

**Performance Level:**
**Student Scaled Score:**



## Scores on Montana Content Standards

In addition to performance levels, CRT results are reported for Montana Content Standards in reading and mathematics. Unlike scaled scores which provide a total performance level score, Montana Content Standard Scores provide more specific information about your child's achievement on the CRT. The chart on the following page shows your child's performance in each area of study within subject areas (Montana Content Standards for reading and math). These results can be used to show your child's relative strengths or weaknesses.

Contact your student's school for more information about the following symbols:
† Student did not complete the assessment.
§ Student took non-standard accommodation.

## Scores on Montana Standards

| Scores on Montana Standards | Points Possible | Student Percentage | State Percentage | Percentage of Points Earned<br>0 25 50 75 100 |
|---|---|---|---|---|
| **Reading Standards** | | | | |
| 1. Students construct meaning as they comprehend, interpret, and respond to what they read. | 25 | | | |
| 2. Students apply a range of skills and strategies to read. | 15 | | | |
| 3. Students set goals, monitor, and evaluate their reading progress. | This standard is not measureable in a statewide assessment. | | | |
| 4. Students select, read, and respond to print and nonprint material for a variety of purposes. | 13 | | | |
| 5. Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences. | 7 | | | |
| **Math Standards** | | | | |
| 1. Problem Solving | 7 | | | |
| 2. Numbers and Operations | 14 | | | |
| 3. Algebra | 6 | | | |
| 4. Geometry | 11 | | | |
| 5. Measurement | 7 | | | |
| 6. Data Analysis, Statistics, and Probability | 12 | | | |
| 7. Patterns, Relations, and Functions | 7 | | | |

Percentage of points earned by students

State percentage of points earned

**MontCAS, Phase 2 CRT**

# Reading
# Roster & Item-Level Report
**Confidential**

Class:
School:
System:

Spring 2005

Grade: 04
Page: 1 of: 2

| | Item Number | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 46 | Scaled Score | Perf. Level |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Standard | 2 | 2 | 1 | 4 | 5 | 2 | 2 | 1 | 4 | 5 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 5 | 1 | 1 | 5 | 2 | 1 | 5 | 4 | 4 | 4 | 2 | 4 | 1 | | |
| | Correct Response | A | B | D | B | C | A | A | D | D | C | B | C | B | D | B | C | A | D | B | B | A | | B | B | A | A | D | C | B | C | D | B | A | | |
| Name | Total Possible Points | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Class Average** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **School Average** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **System Average** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **State Average** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

† Student did not complete the assessment. § Student took non-standard accommodation. ¥ Not in school and/or district for full academic year.

ShellsShells

MontCAS, Phase 2
CRT

# Reading
# Roster & Item-Level Report
Confidential

Class:
School:
System:

Spring 2005

Grade: 04
Page: 2 of: 2

| | Item Number | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | | | | | | | | | | | | | Scaled Score | Perf. Level |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Standard | 2 | 2 | 4 | 4 | 2 | 1 | 1 | 1 | 5 | 1 | 1 | 2 | 4 | 4 | 2 | 5 | 4 | 2 | 4 | 4 | 1 | | | | | | | | | | | | | | |
| | Correct Response | C | D | A | D | C | D | B | B | C | B | A | C | C | D | C | B | B | A | C | D | | | | | | | | | | | | | | | |
| Name | Total Possible Points | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Class Average | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| School Average | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| System Average | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| State Average | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

† Student did not complete the assessment. § Student took non-standard accommodation. ¥ Not in school and/or district for full academic year.

ShellsShells

# MontCAS, Phase 2 CRT

# Mathematics
# Roster & Item-Level Report
## Confidential

Class:
School:
System:

Spring 2005

Grade: 04
Page: 1 of: 2

| | Item Number | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 25 | 26 | 27 | 28 | 29 | 30 | 35 | 36 | 37 | 38 | 39 | Scaled Score | Perf. Level |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Standard | 5 | 2 | 4 | 6 | 5 | 7 | 6 | 3 | 4 | 7 | 7 | 4 | 2 | 4 | 5 | 2 | 4 | 1 | 1 | 2 | 7 | 3 | 6 | 1 | 3 | 1 | 1 | 4 | 6 | 3 | 6 | 4 | 1 | | |
| | Correct Response | D | C | A | B | D | B | C | D | B | C | B | A | C | B | C | A | D | C | A | B | B | A | ▨ | A | D | B | C | A | D | A | C | B | C | | |
| Name | Total Possible Points | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Class Average | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| School Average | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| System Average | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| State Average | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

† Student did not complete the assessment. § Student took non-standard accommodation. ¥ Not in school and/or district for full academic year.

ShellsShells

**MontCAS, Phase 2 CRT**

## Mathematics
## Roster & Item-Level Report
**Confidential**

Class:
School:
System:

Spring 2005

Grade: 04
Page: 2 of: 2

| | Item Number | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | | | | | | | | | Scaled Score | Perf. Level |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Standard | 2 | 5 | 5 | 1 | 6 | 4 | 6 | 7 | 4 | 5 | 4 | 3 | 2 | 6 | 6 | 7 | 3 | 5 | 7 | 4 | 2 | 2 | 2 | 2 | 2 | | | | | | | | | | |
| | Correct Response | A | C | B | C | D | A | A | D | B | D | B | A | C | D | C | B | C | D | A | B | C | | | | | | | | | | | | | | |
| Name | Total Possible Points | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Class Average** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **School Average** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **System Average** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **State Average** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

† Student did not complete the assessment. § Student took non-standard accommodation. ¥ Not in school and/or district for full academic year.

ShellsShells

# MontCAS, Phase 2
# CRT

System:
Grade: 04
Spring 2005

## Reading

## System Summary Report

### I. Distribution of scores

| Perf. Level | Scores | System | | | State | | |
|---|---|---|---|---|---|---|---|
| | | N | % of Students | % of Students in Cat. | N | % of Students | % of Students in Cat. |
| Advanced | 297-300 | | | | | | |
| | 294-296 | | | | | | |
| | 290-293 | | | | | | |
| | 287-289 | | | | | | |
| | 283-286 | | | | | | |
| Proficient | 276-282 | | | | | | |
| | 270-275 | | | | | | |
| | 263-269 | | | | | | |
| | 257-262 | | | | | | |
| | 250-256 | | | | | | |
| Nearing Proficiency | 245-249 | | | | | | |
| | 240-244 | | | | | | |
| | 235-239 | | | | | | |
| | 230-234 | | | | | | |
| | 225-229 | | | | | | |
| Novice | 220-224 | | | | | | |
| | 215-219 | | | | | | |
| | 210-214 | | | | | | |
| | 205-209 | | | | | | |
| | 200-204 | | | | | | |

### II. Subtest results

| | Reading | Points Possible | Average Points Earned | |
|---|---|---|---|---|
| | | | System | State |
| | Total Points | 60 | | |
| Standards | 1. Students construct meaning as they comprehend, interpret, and respond to what they read | 25 | | |
| | 2. Students apply a range of skills and strategies to read | 15 | | |
| | 3. Students set goals, monitor, and evaluate their reading progress | This standard is not measureable in a statewide assessment. | | |
| | 4. Students select, read, and respond to print and nonprint material for a variety of purposes | 13 | | |
| | 5. Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences | 7 | | |

**CRT Performance Level Descriptors**

**Advanced (283-300)**
This level denotes superior performance.

**Proficient (250-282)**
This level denotes solid academic performance for each benchmark. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.

**Nearing Proficiency (225-249)**
This level denotes that the student has partial mastery or prerequisite knowledge and skills fundamental for proficient work at each benchmark.

**Novice (200-224)**
This level denotes that the student is beginning to attain the prerequisite knowledge and skills that are fundamental for work at each benchmark.

# MontCAS, Phase 2 CRT

**Confidential**

**Reading**

**System Summary Report**

System:
Grade: 04
Spring 2005

## III. Results for Subgroups of Students

| Reporting category | System | | | | | State | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | % in N | % in NP | % in P | % in A | N | % in N | % in NP | % in P | % in A |
| **All Students** | | | | | | | | | | |
| Gender | | | | | | | | | | |
| Male | | | | | | | | | | |
| Female | | | | | | | | | | |
| Ethnicity | | | | | | | | | | |
| American Indian | | | | | | | | | | |
| Asian | | | | | | | | | | |
| Hispanic | | | | | | | | | | |
| Black or African American | | | | | | | | | | |
| Native Hawaiian or Other Pacific Islander | | | | | | | | | | |
| White | | | | | | | | | | |
| Significant Cognitive Disability | | | | | | | | | | |
| Special Education | | | | | | | | | | |
| Students with a 504 Plan | | | | | | | | | | |
| Title I (optional) | | | | | | | | | | |
| Tested with Standard Accommodation | | | | | | | | | | |
| Tested with Non-Standard Accommodation | | | | | | | | | | |
| Alternate Assessment | | If a student in your system or school took the CRT-Alternate, please refer to Table III on the System or School CRT-Alternate Summary Report. | | | | | | | | |
| Migrant | | | | | | | | | | |
| Gifted/Talented | | | | | | | | | | |
| LEP/ELL | | | | | | | | | | |
| Former LEP Student | | | | | | | | | | |
| LEP Student Enrolled for First Time in a U.S. School | | Performance levels are not reported for 1st year LEP students. | | | | | | | | |
| Free/Reduced Lunch | | | | | | | | | | |
| Special Education Disability(ies): | | | | | | | | | | |
| Autism | | | | | | | | | | |
| Child with a Disability | | | | | | | | | | |
| Cognitive Delay | | | | | | | | | | |
| Deaf-Blindness Impairment | | | | | | | | | | |
| Deafness | | | | | | | | | | |
| Emotional Disturbance | | | | | | | | | | |
| Hearing Impairment | | | | | | | | | | |
| Learning Disability | | | | | | | | | | |
| Orthopedic Impairment | | | | | | | | | | |
| Other Health Impairment | | | | | | | | | | |
| Speech/Language | | | | | | | | | | |
| Traumatic Brain Injury | | | | | | | | | | |
| Visual Impairment | | | | | | | | | | |

*Less than ten (10) students were assessed.

# MontCAS, Phase 2
# CRT

System:
Grade: 04
Spring 2005

Mathematics

# System Summary Report

## I. Distribution of scores

| Perf. Level | Scores | System | | | State | | |
|---|---|---|---|---|---|---|---|
| | | N | % of Students | % of Students in Cat. | N | % of Students | % of Students in Cat. |
| Advanced | 298-300 | | | | | | |
| | 295-297 | | | | | | |
| | 292-294 | | | | | | |
| | 289-291 | | | | | | |
| | 286-288 | | | | | | |
| Proficient | 279-285 | | | | | | |
| | 272-278 | | | | | | |
| | 264-271 | | | | | | |
| | 257-263 | | | | | | |
| | 250-256 | | | | | | |
| Nearing Proficiency | 245-249 | | | | | | |
| | 240-244 | | | | | | |
| | 235-239 | | | | | | |
| | 230-234 | | | | | | |
| | 225-229 | | | | | | |
| Novice | 220-224 | | | | | | |
| | 215-219 | | | | | | |
| | 210-214 | | | | | | |
| | 205-209 | | | | | | |
| | 200-204 | | | | | | |

## II. Subtest results

| Mathematics | | Points Possible | Average Points Earned | |
|---|---|---|---|---|
| | | | System | State |
| Total Points | | 64 | | |
| Standards | 1. Problem Solving | 7 | | |
| | 2. Numbers and Operations | 14 | | |
| | 3. Algebra | 6 | | |
| | 4. Geometry | 11 | | |
| | 5. Measurement | 7 | | |
| | 6. Data Analysis, Statistics, and Probability | 12 | | |
| | 7. Patterns, Relations, and Functions | 7 | | |

### CRT Performance Level Descriptors

**Advanced (286-300)**
This level denotes superior performance.

**Proficient (250-285)**
This level denotes solid academic performance for each benchmark. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.

**Nearing Proficiency (225-249)**
This level denotes that the student has partial mastery or prerequisite knowledge and skills fundamental for proficient work at each benchmark.

**Novice (200-224)**
This level denotes that the student is beginning to attain the prerequisite knowledge and skills that are fundamental for work at each benchmark.

# MontCAS, Phase 2 CRT

Confidential

**Mathematics**

**System Summary Report**

System:
Grade: 04
Spring 2005

## III. Results for Subgroups of Students

| Reporting category | System | | | | | State | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | % in N | % in NP | % in P | % in A | N | % in N | % in NP | % in P | % in A |
| **All Students** | | | | | | | | | | |
| Gender | | | | | | | | | | |
| Male | | | | | | | | | | |
| Female | | | | | | | | | | |
| Ethnicity | | | | | | | | | | |
| American Indian | | | | | | | | | | |
| Asian | | | | | | | | | | |
| Hispanic | | | | | | | | | | |
| Black or African American | | | | | | | | | | |
| Native Hawaiian or Other Pacific Islander | | | | | | | | | | |
| White | | | | | | | | | | |
| Significant Cognitive Disability | | | | | | | | | | |
| Special Education | | | | | | | | | | |
| Students with a 504 Plan | | | | | | | | | | |
| Title I (optional) | | | | | | | | | | |
| Tested with Standard Accommodation | | | | | | | | | | |
| Tested with Non-Standard Accommodation | | | | | | | | | | |
| Alternate Assessment | | If a student in your system or school took the CRT-Alternate, please refer to Table III on the System or School CRT-Alternate Summary Report. | | | | | | | | |
| Migrant | | | | | | | | | | |
| Gifted/Talented | | | | | | | | | | |
| LEP/ELL | | | | | | | | | | |
| Former LEP Student | | | | | | | | | | |
| LEP Student Enrolled for First Time in a U.S. School | | Performance levels are not reported for 1st year LEP students. | | | | | | | | |
| Free/Reduced Lunch | | | | | | | | | | |
| Special Education Disability(ies): | | | | | | | | | | |
| Autism | | | | | | | | | | |
| Child with a Disability | | | | | | | | | | |
| Cognitive Delay | | | | | | | | | | |
| Deaf-Blindness Impairment | | | | | | | | | | |
| Deafness | | | | | | | | | | |
| Emotional Disturbance | | | | | | | | | | |
| Hearing Impairment | | | | | | | | | | |
| Learning Disability | | | | | | | | | | |
| Orthopedic Impairment | | | | | | | | | | |
| Other Health Impairment | | | | | | | | | | |
| Speech/Language | | | | | | | | | | |
| Traumatic Brain Injury | | | | | | | | | | |
| Visual Impairment | | | | | | | | | | |

*Less than ten (10) students were assessed.

# MontCAS, Phase 2
# CRT

School:
System:
Grade: 04
Spring 2005

**Reading**

## School Summary Report

### I. Distribution of scores

| Perf. Level | Scores | School | | | System | | | State | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | N | % of Students | % of Students in Cat. | N | % of Students | % of Students in Cat. | N | % of Students | % of Students in Cat. |
| Advanced | 297-300 | | | | | | | | | |
| | 294-296 | | | | | | | | | |
| | 290-293 | | | | | | | | | |
| | 287-289 | | | | | | | | | |
| | 283-286 | | | | | | | | | |
| Proficient | 276-282 | | | | | | | | | |
| | 270-275 | | | | | | | | | |
| | 263-269 | | | | | | | | | |
| | 257-262 | | | | | | | | | |
| | 250-256 | | | | | | | | | |
| Nearing Proficiency | 245-249 | | | | | | | | | |
| | 240-244 | | | | | | | | | |
| | 235-239 | | | | | | | | | |
| | 230-234 | | | | | | | | | |
| | 225-229 | | | | | | | | | |
| Novice | 220-224 | | | | | | | | | |
| | 215-219 | | | | | | | | | |
| | 210-214 | | | | | | | | | |
| | 205-209 | | | | | | | | | |
| | 200-204 | | | | | | | | | |

### II. Subtest results

| Reading | | Points Possible | Average Points Earned | | |
|---|---|---|---|---|---|
| | | | School | System | State |
| Total Points | | 60 | | | |
| Standards | 1. Students construct meaning as they comprehend, interpret, and respond to what they read | 25 | | | |
| | 2. Students apply a range of skills and strategies to read | 15 | | | |
| | 3. Students set goals, monitor, and evaluate their reading progress | This standard is not measureable in a statewide assessment. | | | |
| | 4. Students select, read, and respond to print and nonprint material for a variety of purposes | 13 | | | |
| | 5. Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences | 7 | | | |

**CRT Performance Level Descriptors**

**Advanced (283-300)**
This level denotes superior performance.

**Proficient (250-282)**
This level denotes solid academic performance for each benchmark. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.

**Nearing Proficiency (225-249)**
This level denotes that the student has partial mastery or prerequisite knowledge and skills fundamental for proficient work at each benchmark.

**Novice (200-224)**
This level denotes that the student is beginning to attain the prerequisite knowledge and skills that are fundamental for work at each benchmark.

# MontCAS, Phase 2 CRT

**Confidential**

**Reading** **School Summary Report**

School:
System:
Grade: 04
Spring 2005

## III. Results for Subgroups of Students

| Reporting category | School | | | | | System | | | | | State | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | % in N | % in NP | % in P | % in A | N | % in N | % in NP | % in P | % in A | N | % in N | % in NP | % in P | % in A |
| **All Students** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Gender |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Male |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Female |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Ethnicity |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| American Indian |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Asian |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Hispanic |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Black or African American |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Native Hawaiian or Other Pacific Islander |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| White |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Significant Cognitive Disability |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Special Education |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Students with a 504 Plan |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Title I (optional) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Tested with Standard Accommodation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Tested with Non-Standard Accommodation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Alternate Assessment |  | If a student in your system or school took the CRT-Alternate, please refer to Table III on the System or School CRT-Alternate Summary Report. |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Migrant |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Gifted/Talented |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| LEP/ELL |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Former LEP Student |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| LEP Student Enrolled for First Time in a U.S. School |  | Performance levels are not reported for 1st year LEP students. |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Free/Reduced Lunch |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Special Education Disability(ies): |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Autism |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Child with a Disability |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Cognitive Delay |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Deaf-Blindness Impairment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Deafness |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Emotional Disturbance |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Hearing Impairment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Learning Disability |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Orthopedic Impairment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Other Health Impairment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Speech/Language |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Traumatic Brain Injury |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Visual Impairment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

*Less than ten (10) students were assessed.

# MontCAS, Phase 2
# CRT

School:
System:
Grade: 04
Spring 2005

**Mathematics**

# School Summary Report

## I. Distribution of scores

| Perf. Level | Scores | School | | | System | | | State | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | N | % of Students | % of Students in Cat. | N | % of Students | % of Students in Cat. | N | % of Students | % of Students in Cat. |
| Advanced | 298-300 | | | | | | | | | |
| | 295-297 | | | | | | | | | |
| | 292-294 | | | | | | | | | |
| | 289-291 | | | | | | | | | |
| | 286-288 | | | | | | | | | |
| Proficient | 279-285 | | | | | | | | | |
| | 272-278 | | | | | | | | | |
| | 264-271 | | | | | | | | | |
| | 257-263 | | | | | | | | | |
| | 250-256 | | | | | | | | | |
| Nearing Proficiency | 245-249 | | | | | | | | | |
| | 240-244 | | | | | | | | | |
| | 235-239 | | | | | | | | | |
| | 230-234 | | | | | | | | | |
| | 225-229 | | | | | | | | | |
| Novice | 220-224 | | | | | | | | | |
| | 215-219 | | | | | | | | | |
| | 210-214 | | | | | | | | | |
| | 205-209 | | | | | | | | | |
| | 200-204 | | | | | | | | | |

## II. Subtest results

| Mathematics | | Points Possible | Average Points Earned | | |
|---|---|---|---|---|---|
| | | | School | System | State |
| Total Points | | 64 | | | |
| Standards | 1. Problem Solving | 7 | | | |
| | 2. Numbers and Operations | 14 | | | |
| | 3. Algebra | 6 | | | |
| | 4. Geometry | 11 | | | |
| | 5. Measurement | 7 | | | |
| | 6. Data Analysis, Statistics, and Probability | 12 | | | |
| | 7. Patterns, Relations, and Functions | 7 | | | |

### CRT Performance Level Descriptors

**Advanced (286-300)**
This level denotes superior performance.

**Proficient (250-285)**
This level denotes solid academic performance for each benchmark. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.

**Nearing Proficiency (225-249)**
This level denotes that the student has partial mastery or prerequisite knowledge and skills fundamental for proficient work at each benchmark.

**Novice (200-224)**
This level denotes that the student is beginning to attain the prerequisite knowledge and skills that are fundamental for work at each benchmark.

## Confidential

# MontCAS, Phase 2 CRT

**Mathematics**

**School Summary Report**

School:
System:
Grade: 04
Spring 2005

## III. Results for Subgroups of Students

| Reporting category | School | | | | | System | | | | | State | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | % in N | % in NP | % in P | % in A | N | % in N | % in NP | % in P | % in A | N | % in N | % in NP | % in P | % in A |
| **All Students** | | | | | | | | | | | | | | | |
| Gender | | | | | | | | | | | | | | | |
| Male | | | | | | | | | | | | | | | |
| Female | | | | | | | | | | | | | | | |
| Ethnicity | | | | | | | | | | | | | | | |
| American Indian | | | | | | | | | | | | | | | |
| Asian | | | | | | | | | | | | | | | |
| Hispanic | | | | | | | | | | | | | | | |
| Black or African American | | | | | | | | | | | | | | | |
| Native Hawaiian or Other Pacific Islander | | | | | | | | | | | | | | | |
| White | | | | | | | | | | | | | | | |
| Significant Cognitive Disability | | | | | | | | | | | | | | | |
| Special Education | | | | | | | | | | | | | | | |
| Students with a 504 Plan | | | | | | | | | | | | | | | |
| Title I (optional) | | | | | | | | | | | | | | | |
| Tested with Standard Accommodation | | | | | | | | | | | | | | | |
| Tested with Non-Standard Accommodation | | | | | | | | | | | | | | | |
| Alternate Assessment | | If a student in your system or school took the CRT-Alternate, please refer to Table III on the System or School CRT-Alternate Summary Report. | | | | | | | | | | | | | |
| Migrant | | | | | | | | | | | | | | | |
| Gifted/Talented | | | | | | | | | | | | | | | |
| LEP/ELL | | | | | | | | | | | | | | | |
| Former LEP Student | | | | | | | | | | | | | | | |
| LEP Student Enrolled for First Time in a U.S. School | | Performance levels are not reported for 1st year LEP students. | | | | | | | | | | | | | |
| Free/Reduced Lunch | | | | | | | | | | | | | | | |
| Special Education Disability(ies): | | | | | | | | | | | | | | | |
| Autism | | | | | | | | | | | | | | | |
| Child with a Disability | | | | | | | | | | | | | | | |
| Cognitive Delay | | | | | | | | | | | | | | | |
| Deaf-Blindness Impairment | | | | | | | | | | | | | | | |
| Deafness | | | | | | | | | | | | | | | |
| Emotional Disturbance | | | | | | | | | | | | | | | |
| Hearing Impairment | | | | | | | | | | | | | | | |
| Learning Disability | | | | | | | | | | | | | | | |
| Orthopedic Impairment | | | | | | | | | | | | | | | |
| Other Health Impairment | | | | | | | | | | | | | | | |
| Speech/Language | | | | | | | | | | | | | | | |
| Traumatic Brain Injury | | | | | | | | | | | | | | | |
| Visual Impairment | | | | | | | | | | | | | | | |

*Less than ten (10) students were assessed.

ShellsShells

# APPENDIX F: REPORTING DECISION RULES

**MontCAS, Phase 2 CRT Spring 04-05**
**Decision Rules**

| Participation | | | Relationship w/ Data File Layouts | Impact on Analyses | Impact on Student report | Impact on School/System/State reports | Impact on Student Roster and I-Analyze | Impact on student level data Excel files for System CD’s | Impact on student level data Excel files for State CD |
|---|---|---|---|---|---|---|---|---|---|
| Number of Students (“N”) | 1 | Number of students included in state aggregation | NA | N=total number of students with 2 or more responses minus students tested at a private accredited school (PRAS) minus students tested in a non-accredited Title I private school (PRNONST) minus foreign exchange (FXS) students minus students not enrolled (SNE) minus student enrolled part-time (PSNE) minus students tested at a private non-accredited school (PRNAS) minus LEP student enrolled first time in U.S. school | | | | | |
| No class header provided | 2 | No class indicators provided | Tfname=’ ‘ and Tlname=’ ‘ | Class aggregations calculated are actually school level. | No impact | No impact | Report produced | No impact | No impact |
| Number of Students for Reporting | 3 | Schools (Systems) has **less than 10 included** students in both content areas | NA | | No impact | School/system report Produced. Page 2: For each category numbers will be suppressed if number of included students less than ten. The N-size is always reported. Footnote *’Less than 10 students were assessed” | No Impact | No Impact | No Impact |
| Student Names Not Provided | 4 | No student barcode label and no name bubbled on answer sheet | Lname, Fname | No Impact on analyses.<br>Student included in DP report to systems.<br>Student counted in N. | No Impact<br>Student name is “Name Not Provided” | Student included based on inclusion rules stated in this document. | Student included based on inclusion rules stated in this document. | Student included based on inclusion rules stated in this document. | Student included based on inclusion rules stated in this document. |

| Participation | | | Relationship w/ Data File Layouts | Impact on Analyses | Impact on Student report | Impact on School/System/State reports | Impact on Student Roster and I-Analyze | Impact on student level data Excel files for System CD's | Impact on student level data Excel files for State CD |
|---|---|---|---|---|---|---|---|---|---|
| Form Number Not Coded | 5 | DP codes as Form 1, only common items scored | Form | Student counted in N | No Impact | Student Included | No Impact | No Impact | No impact. |
| Tested but Fewer than 2 of the answers marked | 6 | Student answered fewer than 2 of the common MC questions | All common items | Student not counted in N; student excluded from item analysis | Score given with a footnote (†): "Student did not complete the assessment" | Student not Included | Score given with a footnote (†): "Student did not complete the assessment" | Student Included | No Impact |
| Tested with Standard Accommodations | 7 | Student requires an accommodation(s) by content area | Any REASA1-REASA28 bubbled and MATSA1-MATSA28 | If one or more standard accommodations (#1-28) are coded, student is counted as Tested with Standard Accommodation(s) | No Impact | Counted as Tested with Standard Accommodation(s) | No Impact | No Impact | No Impact |
| Tested with Non-standard Accommodations | 8 | Student requires a non-standard accommodation(s) by content area | Any REANSA29-REANSA32 bubbled and MATNSA29-MATNSA32 | If one or more non-standard accommodations (#29-32) are coded, student is counted as Tested with Non-standard Accommodation(s) and will receive a performance level of "NOVICE" and lowest possible scaled score in content area(s) where non-standard accommodations were coded. | Student report will indicate raw score with an (§) and a footnote stating that the student took a non-standard accommodation. The scaled score is the scaled score associated with the earned raw score | Student will be given a performance level of "NOVICE" and be included in student counts in content area(s) where non-standard accommodations were coded. | Student record will indicate raw score with an (§) stating that the student took a non-standard accommodation. | Student included with earned scaled score and performance level | Student will score "NOVICE" and be included in student counts in content area(s) where non-standard accommodations were coded. |

| Participation | | | Relationship w/ Data File Layouts | Impact on Analyses | Impact on Student report | Impact on School/System/State reports | Impact on Student Roster and I-Analyze | Impact on student level data Excel files for System CD’s | Impact on student level data Excel files for State CD |
|---|---|---|---|---|---|---|---|---|---|
| Program Information | 9 | Student is identified as participating in an identified program. | SE=’1’ or Plan504=’1’ or Migrant=’1’ or GT=’1’ or LEP=’1’ or Lunch=’1’ or TM or TR Disab=’1’ | If one or more Program Information codes are bubbled, student is counted as a program participant **LEP students do not included LEP students enrolled first time in U.S. school.** | No Impact | Reported on school & system Reporting Category reports. All numbers except the N-size are suppressed if N-size less than 10. Footnote *’Less than 10 students were assessed.’ | No Impact | No Impact | No Impact |
| Special Education- not optional. Can have more than one bubbled for a student | 10 | Student is has an identified disability under IDEA -97. | AU=’1’,CW=’1’,CD=’1’,DB=’1’,DE=’1’, ED=’1’,HI=’1’,LD=’1’,OI=’1’,OH=’1’,SL=’1’,TB=’1’,VI=’1’ | Student is counted in their respective disability group on page 2 of summary reports. | No Impact | Student is counted in their respective disability group on page 2 of summary reports. All numbers except the N-size are suppressed if N-size less than 10.Footnote * ‘Less than 10 students were assessed.’ | No Impact | No Impact | No Impact |
| First year LEP student | 11 | Student is identified as being a first year LEP | Exclusions=’1’ | Student is excluded from all aggregations for both content areas. | Student receives report. Student does not receive scaled score or performance level for reading. Performance Level=’LEP’ on report for reading. Student receives earned score in Math. | Student is excluded from aggregations**. Student is counted in First year LEP student enrolled first time in U.S. school.** | Student is included. Student’s scaled score is blank. Student’s performance level =’LEP’ for Reading. If student took the Reading test the responses are shown. The student included in Math with earned scores and responses shown. | Student is included | Student is included |
| Foreign Exchange Student (FXS) | 12 | Student is identified as a foreign exchange student | Exclusions=’2’ | Student is not included in any school/system/state aggregations. | Student receives report. | Not included on Reports | Not Included on Reports | Students are not included on System CD | Included on State CD; identified as FXS |
| Student Not Enrolled (SNE) | 13 | Student is identified as not enrolled in an accredited public | Exclusions=’3’ | Student is not included in any school/system/state aggregations | Student receives report. | Not Included on Reports | Not Included on Reports | Students are not included on System CD | Not included on State CD |

| Participation | | | Relationship w/ Data File Layouts | Impact on Analyses | Impact on Student report | Impact on School/System/State reports | Impact on Student Roster and I-Analyze | Impact on student level data Excel files for System CD's | Impact on student level data Excel files for State CD |
|---|---|---|---|---|---|---|---|---|---|
| (Homeschooled) | | school. | | | | | | | |
| Private Accredited School (PRAS) | 14 | Student is identified as testing at a private accredited school | Exclusions='5' | Student is not included in any state aggregations | Student receives report. | School report produced. System report produced. They are their own system. | Report produced | Students are included on System CD | Included on State CD; identified as PRAS |
| Private Non-accredited Title I School (PRNONST) | 15 | Student is identified as testing in a non-accredited Title I school | Exclusions='7' | Student is not included in any state aggregations | Student receives report. | School report produced. System report produced. They are their own system. | Report produced | Students are included on System CD | Included on State CD; identified as PRNONST |
| Private Non-Accredited School (PRNAS) | 16 | Student is identified as testing in a non-accredited school. | Exclusions='6' | Student is not included in state aggregations | Student receives report. | School report produced. System report produced. They are their own system. | Report produced. | Students are included on System CD. | Included on state CD; identified as PRNAS |
| Student enrolled part-time (<180 hours) (PSNE) | 17 | Student is identified as enrolled part-time | Exclusions='4' | Student is not included in any school/system/state aggregations | Student receives report. | Student not included | Not included on reports. | Students are not included in system CD. | Student included on state CD; identified as PSNE |
| Former LEP | 18 | Former LEP student | FLEP='1' | Student is included in all aggregations | Student receives report | Student included. Counted in category on page 2.All numbers except N-size are suppressed if N-size less than 10. | Student included | Student included | Student included |

| Participation | | | Relationship w/ Data File Layouts | Impact on Analyses | Impact on Student report | Impact on School/System/State reports | Impact on Student Roster and I-Analyze | Impact on student level data Excel files for System CD's | Impact on student level data Excel files for State CD |
|---|---|---|---|---|---|---|---|---|---|
| LEP student currently receiving Title III Services(not first year LEP) | 19 | LEP student currently receiving Title III Services | Title3='1' | Student is included in all aggregations | Student receives report | Student included.<br>**Student counted with LEP student category on page 2 of summary reports.** | Student included | Student included | Student included |
| 1st year LEP enrolled first time in U.S school | 20 | LEP student enrolled for first time in U.S School | LEPFirst ='1' | Student is excluded from all aggregations for both content areas. | Student receives report. Student does not receive scaled score or performance level for reading. Performance Level='LEP' on report for reading. Student receives earned score in Math. | Student is excluded from aggregations<br>Included in the count of 1st year LEP enrolled first time in U.S school students on Page 2 of summary reports. Only N-size is reported. The rest of the line is covered with a watermark on Reading reports. | Student is included. Student's scaled score is blank. Student's performance level ='LEP' for Reading. If student took the Reading test the responses are shown. The student included in Math with earned scores and responses shown. | Student is included | Student is included |
| Participation Information (NSAY & NDAY) | 21 | Student participated in CRT but has not been a student in school or district for entire academic year. | NA | Student is included in participation. If student is marked as NSAY only then student is not included in school aggregations. If student is marked as NDAY then student is not included in either school or district aggregations. | No impact. | If student is marked as NSAY only then student is not included in school data. If student is marked as NDAY then student is not included in school or district data. | If student is NSAY or NDAY student is included on roster with footnote(¥)<br>"Not in school and/or district full academic year." Student excluded from school (if NSAY or NDAY) and/or district (if NDAY) aggregations. | No Impact | No Impact |
| Alternate Assessment Student | 22 | Student participated through alternate assessment this year | Alt='1' | Student is excluded from CRT aggregations. | Student receives CRT-Alternate student report. | Student included in CRT-Alternate aggregations. Student included in count of alternate students on page 2 of CRT summary reports. Only N-size is reported. The rest of the | Student not included in I-Analyze. Student included (unless otherwise excluded based on CRT-Alternate decision rules) on CRT- | Student included in CRT-Alternate system CDs. | Student included in CRT-Alternate state CD. |

| Participation | | | Relationship w/ Data File Layouts | Impact on Analyses | Impact on Student report | Impact on School/System/State reports | Impact on Student Roster and I-Analyze | Impact on student level data Excel files for System CD's | Impact on student level data Excel files for State CD |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | line is covered with a watermark. | Alternate Roster | | |
| Student has bubbled SNE and PSNE | 23 | Student has bubbled both not enrolled and part-time | Exclusions='3' and Exclusions='4' | Student is not included in any school/system/state aggregations | Student receives report. | Student is not included | Student is not included | Student is not included | Student is not included |
| Student is SNE and enrolled part-time(PSNE) in a private school | 24 | Student has bubbled not enrolled and part-time and a private school | Exclusions='3' and '4' and either (5,6, or 7) | Student is not included in any school/system/state aggregations | Student receives report. | Student is not included | Student is not included | Student is not included | Student is not included. |

## Additional Rules:

**1. Only common items are used to calculate scores.**

**Schools: 839**
**Systems: 281**

**Scores:**

**Reading Subtest:** **Raw score is number of correct responses to common items. Total possible is:**

**Grade 4: 60 score points**
**Grade 8: 60 score points**
**Grade 10: 65 score points**

**Math Subtest: Raw score is number of correct responses to common items. Total possible is:**

**Grade 4: 66 score points**
**Grade 8: 66 score points**
**Grade 10: 71 score points**